Learning Conservation Laws in Unknown Quantum Dynamics

Yongtao Zhan[1,2,*] Andreas Elben,[1,3,†] Hsin-Yuan Huang,[1,4,‡] and Yu Tong[1,§]

[1]*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA*

[2]*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA*

[3]*Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, California 91125, USA*

[4]*Google Quantum AI, Venice, California 90291, USA*

We present a learning algorithm for discovering conservation laws given as sums of geometrically local observables in quantum dynamics. This includes conserved quantities that arise from local and global symmetries in closed and open quantum many-body systems. The algorithm combines the classical shadow formalism for estimating expectation values of observable and data analysis techniques based on singular value decompositions and robust polynomial interpolation to discover all such conservation laws in unknown quantum dynamics with rigorous performance guarantees. Our method can be directly realized in quantum experiments, which we illustrate with numerical simulations, using closed and open quantum system dynamics in a $\mathbb{Z}_2$ gauge theory and in many-body localized spin chains.

I. INTRODUCTION

Machine learning (ML) is playing an increasingly important role in physical sciences [1]. The ability of ML to recognize patterns in data greatly facilitates the data-driven approach to scientific research, where scientific discoveries are achieved by analyzing experimental data. Recently, many works have been done to discover physical laws with ML models [2–14]. Following this line, several works attempt to learn conservation laws in classical mechanical systems [15,16]. The ML models in these works can successfully discover conserved quantities in simple classical systems, such as energy conservation, angular momentum conservation, and momentum conservation in two-body gravitational systems.

The conservation law, or the integral of motion, is also an important concept in quantum mechanics. There are usually many global conservation laws in quantum dynamics, such as the eigenstate projection operators,

when the dynamics are governed by a Hamiltonian. However, these do not imply any special dynamical properties. In the quantum setting, the physically relevant conservation laws are those with locality structure [17]. Such local integrals of motion [18–21] underlie, for instance, the absence of thermalization and transport in certain quantum systems—in contrast to ergodic systems, which typically conserve only a few globally supported quantities such as the total energy or number of particles. Thus, local integrals of motion are central for our understanding of phenomena such as many-body localization (MBL) [22–28] and Hilbert space fragmentation [29].

In this work, we consider a broad class of conservation laws with conserved quantities given by sums of geometrically local observables. These conservation laws can be geometrically localized or have support across the entire system. We propose an algorithm for discovering all such conservation laws in arbitrary quantum dynamical systems. We consider different types of quantum dynamics in which a quantum state $\rho(t)$ evolves with time, and its evolution is described by the von Neumann equation under a Hamiltonian H or a Lindblad master equation under a Lindbladian $\mathcal{L}$. Our algorithm is general enough to cover Hamiltonians that change over time, for instance, periodically, as in Floquet systems, and find observables that are conserved between periods. The conservation laws we consider could be state dependent, i.e., the observables are conserved for a certain subset of initial states. Such observables can be more difficult to find than those that are

---

*yzhan@caltech.edu

†aelben@caltech.edu

‡hsinyuan@caltech.edu

§yutong@caltech.edu

conserved for all states because they will be overlooked if only information from the Hamiltonian or the Lindbladian is used.

Our algorithm combines classical shadow formalism [30] and several data analysis techniques. The algorithm uses classical shadows to estimate the expectation values of many Pauli observables at multiple times during the quantum dynamics based on a limited number of randomized measurements [30,31]. After estimating the expectation values, the algorithm performs singular value decomposition (SVD) on a data matrix and gets the low-dimensional manifold corresponding to the conservation laws. We rigorously prove that the algorithm can efficiently learn all conserved quantities that are sums of geometrically local observables in quantum dynamics. Although some previous papers are using numerical techniques to find such conservation laws based on known Hamiltonians [32–36], our method can be directly applied to experiments to find conservation laws in arbitrary unknown quantum dynamics. Our method also comes with theoretical guarantees that the sample and computational complexities are both polynomial in the system size and precision. Note that, for Hamiltonian dynamics, one could also obtain conservation laws by first learning the Hamiltonian [37–53]. However, Hamiltonian learning protocols only work efficiently when there are some known sparsity or locality constraints on the Hamiltonian. For quantum dynamics without such sparsity structure, prior works involve exponential sample complexity or classical post-processing cost. In comparison, our methods are not subject to these constraints. In addition, they allow for the learning of state-dependent conservation laws and are not restricted to Hamiltonian dynamics.

We perform numerical experiments illustrating the learning of conserved quantities in closed and open system dynamics in a $\mathbb{Z}_2$ gauge theory, arising from local and global symmetries. In addition, we demonstrate the learning of local, approximate conservation laws in a one-dimensional *XXZ* chain with local disorder. Here, a sharp increase in the number of local conserved quantities takes place at a certain disorder strength, which we successfully observe from the result of our algorithm.

Discovering conservation laws in quantum systems is a fundamental problem and has also been studied in prior works. To our knowledge, this is the first work to provide general rigorous guarantees for learning and testing conservation laws in quantum experiments with unknown dynamics, by employing the randomized measurement toolbox [31] and the classical shadow formalism [30]. While Bentsen *et al.* [36] proposed a similar algorithm to construct local conserved quantities in quantum dynamics, they used it in classical simulations of known time-independent Hamiltonian dynamics to study integrability. Shtanko *et al.* [54] conducted physical experiments to obtain conserved quantities that are supported in local regions, which excludes conserved quantities supported on the entire system, such as total magnetization. In contrast, our protocol can uncover conservation laws supported locally and globally. Furthermore, Shtanko *et al.* [54] considered quantities that are conserved at discrete points in time, whereas we can also deal with the continuous-time scenario through robust polynomial interpolation.

II. ALGORITHM DESCRIPTION

The goal of our algorithm is to find all conserved quantities that are linear combinations of Pauli operators supported on $k = \mathcal{O}(1)$ adjacent qubits, in an unknown quantum dynamical system with experimentally feasible measurements. The Hamiltonian or Lindbladian that governs the quantum dynamics is completely unknown and need not be local, but we can control it to evolve for a time of our choice and perform randomized single-qubit measurements. Throughout this work, we use $\rho(t)$ to denote the time-evolved state and $O(t)$ to denote the time-evolved observable in the Heisenberg picture.

A. Classical shadows

The classical shadow formalism [30] was proposed to efficiently predict local observables with experimentally feasible randomized measurements [31]. To be specific, it was shown that one can predict M arbitrary linear target functions $\text{Tr}(O_1\rho), \ldots, \text{Tr}(O_M\rho)$ up to an additive error ϵ with only $\mathcal{O}(B\log(M)/\epsilon^2)$ measurements, where B is the upper bound of the shadow norm defined in Ref. [30]. This result implies that a limited number of measurements are enough to predict the expectation values of a large number of observables. Making use of this property, we can predict all k-local Pauli observables with a limited number of random measurements.

The classical shadow formalism is summarized as follows. We approximate an N-qubit quantum state by performing randomized single-qubit Pauli measurements on N_s copies of ρ. That is, we project each qubit to one of three Pauli bases X, Y, Z and get a product state composed of the six basis states $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i+\rangle, |i-\rangle\}$. Performing one randomized measurement gives us such a product state, which can be stored in classical memory with an N-element array. After performing such measurements on N_s copies of states, we get NN_s single-qubit measurement results, which we can use to construct an approximation of the unknown state ρ:

$$\hat{\rho} = \frac{1}{N_s} \sum_{n_s=1}^{N_s} \hat{\rho}_1^{(n_s)} \otimes \cdots \otimes \hat{\rho}_N^{(n_s)}. \tag{1}$$

Here $\hat{\rho}_i^{(n_s)} = 3|s_i^{(n_s)}\rangle\langle s_i^{(n_s)}| - \mathbb{I}$ and $s_i^{(n_s)}$ is the outcome of qubit i in the n_sth randomized measurement. Equation (1), in principle, allows us to fully recover the density

matrix ρ, in the sense that $\mathbb{E}[\hat{\rho}] = \rho$, where we take the expectation value of many random unitaries and projective measurements [31]. However, this requires an exponentially large number of copies of states $N_s = \mathcal{O}(\exp(N))$. In contrast, $N_s = \mathcal{O}(3^r \log(N)/\epsilon^2)$ is enough to provide an ϵ-accurate approximation of all r-body reduced density matrices, allowing us to estimate expectation values of r-local observables. We emphasize that this estimation can be made robust against errors in the application of the random unitaries and readout errors [55–57].

B. Learning conservation laws

Our algorithm makes use of classical shadow formalism to evaluate the expectation values of all geometrically k-local Pauli observables with randomized measurements (the measurement results can be used to estimate all k-local Pauli observables, but we only focus on the geometrically local ones), and then postprocess the measurement data to identify the conserved quantities. This is a nontrivial task because there are uncountably many linear combinations of these Pauli observables that can possibly be conserved. Here we propose a method to efficiently narrow down the range of conserved quantities to look for.

We consider a quantum system on a D-dimensional lattice, with each site containing a qubit. We denote all geometrically k-local Pauli operators by P_i, $i = 1, 2, \ldots, N_P$, with $N_P \leq \mathcal{O}(N)$. We then look for conserved quantities of the form

$$O = \sum_i c_i P_i. \tag{2}$$

In other words, the sequence $\{P_i\}$ forms a basis of the subspace in which we search for conserved quantities.

The expectation values of P_i at time t_j, $j = 1, 2, \ldots, N_T$, form a data matrix of size $N_P \times N_T$ ($N_T = \mathcal{O}(N_P)$), which we denote by X. Its elements are

$$X_{ij} = \langle P_i(t_j) \rangle. \tag{3}$$

Our algorithm is built upon the following observation (see also Refs. [36,47]): every conserved quantity of form (2) lies in the null space of a matrix $W^\top$, where its transpose matrix W is defined through

$$W_{ij} = \langle P_i(t_j) \rangle - \frac{1}{N_T} \sum_{j'} \langle P_i(t_{j'}) \rangle. \tag{4}$$

This is because, if operator O defined in Eq. (2) is conserved then the $\sum_i c_i \langle P_i(t_j) \rangle$ are equal for all j, and they

are thus all equal to the average. Consequently,

$$\sum_i c_i \langle P_i(t_j) \rangle = \frac{1}{N_T} \sum_{ij'} c_i \langle P_i(t_{j'}) \rangle. \tag{5}$$

By Eq. (4) we then have $W^\top \vec{c} = 0$, where $\vec{c} = (c_1, c_2, \ldots, c_{N_P})$ is the vector formed by the coefficients in O.

From the above analysis, we can see that all conservation laws that are linear combinations of geometrically k-local terms must correspond to a vector in the null space of $W^\top$, and, consequently, we can find all of them by examining this null space. At the same time, the dimension of this null space yields the number of independent conservation laws. Each singular value λ of W describes how much its corresponding operator expectation value changes over time [36]. More precisely, let $u = (u_1, u_2, \ldots, u_{N_P})$ be the left singular vector corresponding to λ, and let $O = \sum_i u_i P_i$; then

$$\lambda^2 = \sum_j \left| \langle O(t_j) \rangle - \frac{1}{N_T} \sum_{j'} \langle O(t_{j'}) \rangle \right|^2. \tag{6}$$

Because of the inevitable shot noise, the matrix $W^\top$ we get from data will most likely not have a nontrivial null space. Therefore, instead of looking at the null space, we look at the subspace spanned by the left singular vectors of W corresponding to singular values below a truncation threshold ϵ, which serves as a precision parameter. These singular vectors are readily obtainable by performing SVD on our approximation of W based on finitely many samples (see also Ref. [47] for a similar procedure in the context of Hamiltonian learning). The number of such singular values provides an upper bound of the number of independent conserved quantities, which we prove later. The computational cost of performing SVD on the data matrix is $\mathcal{O}(N^3)$.

So far, we have mainly been concerned with conserved quantities that are specific to a single initial state. We may also learn conserved quantities for a distribution $\mathcal{D}$ of initial states in a similar way. In this scenario, we not only sample times t_j, but also the initial states ρ_k from $\mathcal{D}$ independently, for $k = 1, 2, \ldots, N_I$. The data matrix X is constructed to have N_P rows and $N_T N_I$ columns, consisting of entries $X_{i,jk} = \text{tr}[P_i(t_j)\rho_k]$, where j and k together index the columns. Matrix W is similarly modified to be $W_{i,jk} = X_{i,jk} - N_T^{-1} \sum_{j'} X_{i,j'k}$.

C. Testing conservation laws

The above procedure gives us candidates for conservation laws. However, it is not guaranteed that the quantities we get are indeed conserved, and therefore we need to test the candidates. Testing a finite group symmetry has been

considered by LaBorde and Wilde [58], but their algorithm requires implementing the group action on a quantum computer, whereas we want to keep our procedure to only single-qubit operations. There are two problems that we need to overcome. The first is that in the above we only look at a discrete set of times t_j, and cannot rule out the possibility that some quantity be conserved at these discrete times, but not conserved at other times. The second is that we cannot hope to tell if a quantity is exactly conserved because of the presence of shot noise. Consequently, we formalize the problem into a hypothesis testing problem. We first define how far a quantity deviates from its average up to time T by

$$d(O, \rho) = \max_{t \in [0,T]} \left| \mathrm{tr}[\rho O(t)] - \frac{1}{T} \int_0^T \mathrm{tr}[\rho O(s)] \mathrm{d}s \right|. \quad (7)$$

With $d(O, \rho)$ we introduce the two hypotheses that we want to distinguish:

$$\mathbb{E}_{\rho \sim \mathcal{D}}[d(O, \rho)] = 0 \quad \text{or} \quad \mathbb{E}_{\rho \sim \mathcal{D}}[d(O, \rho)] \geq \epsilon \quad (8)$$

for every candidate O that comes from the learning procedure. The classical shadow technique enables us to process all the O in parallel.

For a fixed $\rho \sim \mathcal{D}$, we compute the maximal deviation of observable O from its time average using robust polynomial interpolation [59]. We first randomly sample the discrete times t_j, and then perform robust polynomial interpolation to obtain values for $\langle O(t) \rangle$ at all times $t \in [0, T]$. Then we can directly compute the maximal deviation from the time average. This enables us to compute $d(O, \rho)$ with high confidence level. Note that the above discussion is for continuous t. If we want to test conservation laws for discrete t, such as for a Floquet system, then the problem comes strictly easier, as interpolation will not be needed. The ensemble average $\mathbb{E}_{\rho \sim \mathcal{D}} d(O, \rho)$ can be computed from finitely many samples of ρ, thus enabling us to solve the hypothesis testing problem in Eq. (8).

III. RIGOROUS GUARANTEES

As noted previously, the estimates for $\mathrm{tr}[P_i(t_j)\rho_k]$ necessarily involve shot noise, i.e., the noise as a result of using measurement outcomes from only a finite number of experiments. For different observables, the shot noise can also be correlated when we measure multiple observables in the same experiment. We then analyze how the noise impacts the result we get, taking into account the possible correlation between shot noise on different observables. First, we analyze that in the learning algorithm based on finding the null space; the algorithm still yields an upper bound of the number of conserved quantities even when shot noise is present.

We denote the number of independent conserved quantities by N_c, and the dimension of the null space of $W^\top$

by D_{null}. It is guaranteed that $N_c \leq D_{\mathrm{null}}$ because a quantity that is conserved at all times must also be conserved at discrete times t_j and for the sampled states ρ_k. We compute W from the data matrix X, but, in practice, we do not directly have access to X, but can only obtain its noisy estimate $\hat{X}$, which leads to a noisy estimate for W that we denote by $\hat{W}$. The $\hat{W}$ estimate is almost surely full rank due to the effect of the noise. We define $E = \hat{X} - X$, and this is the matrix containing all the entrywise errors. Matrix W is perturbed similarly, and the errors can be collected into a matrix whose spectral norm is at most $\|E\|$. Consequently, we cannot directly estimate D_{null}. As discussed before, instead we look at the number of singular values of $\hat{W}$ that are below a threshold ϵ, which we denote by $\hat{D}_{\mathrm{null}}$. For $\hat{D}_{\mathrm{null}}$, we have the following theorem.

Theorem 1. With $\widetilde{\mathcal{O}}(N^3 \epsilon^{-2} \log(\delta^{-1}))$ samples [60], we can compute an integer $\hat{D}_{\mathrm{null}}^{\mathrm{median}}$ satisfying $N_c \leq \hat{D}_{\mathrm{null}}^{\mathrm{median}}$, where N_c is the number of conserved quantities, with probability at least $1 - \delta$. In particular, when the quantum system has constant correlation length, the sample complexity can be reduced to $\widetilde{\mathcal{O}}(N^2 \epsilon^{-2} \log(\delta^{-1}))$.

Here $\hat{D}_{\mathrm{null}}^{\mathrm{median}}$ is the median of $\mathcal{O}(\log(\delta^{-1}))$ independent estimates of $\mathcal{D}_{\mathrm{null}}$ obtained from independent experiments. In the theorem, N_P is the number of geometrically local Pauli terms and is therefore $\mathcal{O}(N)$. Here, $N_I N_T \geq N_P$ is necessary for avoiding numerical artifacts that do not correspond to any conservation laws, but arise due to the rank deficiency in the data matrix. In the numerical results in Sec. IV below we find that $N_I N_T = 2N_P$ is enough for getting the correct estimate, and, as a result, we choose $N_I N_T = \mathcal{O}(N)$. More general choices of N_T, N_I, and N_P are considered in Theorems 4 and 5 of Appendix A. For the proof of this theorem, we refer the reader to Appendix A, in which we bound singular value perturbation that comes from shot noise. From its definition, we can see that $\hat{D}_{\mathrm{null}}$ is a decreasing function of ϵ, and, consequently, for smaller ϵ, we have a tighter upper bound for D_{null} and the number of conserved quantities N_c. If we keep the singular value perturbation below ϵ then the D_{null} 0-singular values of W will still be below ϵ after perturbation, thus ensuring that $D_{\mathrm{null}} \leq \hat{D}_{\mathrm{null}}^{\mathrm{median}}$. This will require more samples as ϵ decreases, as can be seen from Theorem 1.

We can also guarantee that by collecting all the left singular vectors of $\hat{W}$ corresponding to singular values below the threshold ϵ, we have all the conservation laws approximately contained in the span. More precisely, each conservation law, when expressed as a norm-1 vector, will have an overlap with the subspace spanned by these singular vectors, and this overlap is lower bounded by $\sqrt{1 - \|E\|^2/\epsilon^2}$. Therefore, when $\|E\| \ll \epsilon$, we have an accurate description of all conservation laws. For a detailed proof of this bound, see Appendix B.

Next, we provide guarantees that the candidates for conserved quantities from the learning algorithm can be efficiently verified using the procedure described previously.

First, we consider the scenario where the conservation law is specific to a single initial state, i.e., the distribution $\mathcal{D}$ is completely concentrated on ρ.

Theorem 2. We assume that $\mathcal{D}$ is concentrated on a single ρ. Let $f_i(t) = \text{Tr}[\rho(t)O_i]$ for $i = 1, 2, \ldots, \chi$. We further assume that $|d^\ell f_i(t)/dt^\ell| \leq \mathcal{O}(\Gamma^\ell \ell!)$ for all $\ell \geq 1$. Then, for $T > 0$, we can distinguish between the two hypotheses in Eq. (8) for each i with probability at least $1 - \delta$ using $\widetilde{\mathcal{O}}(\Gamma T \epsilon^{-2} \log(\delta^{-1}) \max_i \|O_i\|_{\text{shadow}}^2)$ samples. Here, $\|O_i\|_{\text{shadow}}$ is the shadow norm of O_i, as defined in Ref. [30].

We refer the reader to Appendix C for a detailed proof. We note that $|d^\ell f_i(t)/dt^\ell| \leq \mathcal{O}(\Gamma^\ell \ell!)$ is a very reasonable assumption to make. In addition, for (suitably normalized) observables O_i that are sums of k-local Pauli operators (as obtained from our learning algorithm), we have $\|O_i\|_{\text{shadow}} \leq 3^k$. We show in Appendix F that this assumption holds with $\Gamma = \mathcal{O}(1)$ when the dynamics is described by the von Neumann equation, and the Hamiltonian satisfies certain conditions. These Hamiltonians include geometrically local Hamiltonians and certain power-law interaction Hamiltonians. Without such an assumption, we can also choose $\Gamma = \|H\|$ and then this inequality holds for all Hamiltonians. An extension to the Lindbladian case is straightforward.

Next, we consider a generic initial state distribution $\mathcal{D}$. In this scenario, we can sample ρ_k, $k = 1, 2, \ldots, N_I$, from the distribution $\mathcal{D}$, and test if the observables $O_1, O_2, \ldots, O_\chi$ are conserved for the sampled initial states. This naturally leads to the question of whether we can generalize the testing results for the sampled initial states to the entire distribution. The above involves generalization errors of the form

$$\left| \mathbb{E}_{\rho \sim \mathcal{D}} d(O_i, \rho) - \frac{1}{N_I} \sum_{k=1}^{N_I} d(O_i, \rho_k) \right|. \quad (9)$$

In Appendix D, we show that we can use $N_I = \mathcal{O}(\epsilon^{-2} \log(\chi \delta^{-1}) \max_i \|O_i\|^2)$ to ensure that the generalization errors for all observables are below $\epsilon/4$ with probability at least $1 - \delta/2$. For each sampled ρ_k, we need $\widetilde{\mathcal{O}}(\Gamma T \epsilon^{-2} \log(\delta^{-1}) \max_i \|O_i\|_{\text{shadow}}^2)$ samples to compute $d(O_i, \rho_k)$, $i = 1, 2, \ldots, \chi$, according to Theorem 7 in Appendix C 2, which multiplied by N_I yields the total sample complexity for estimating $\mathbb{E}_{\rho \sim \mathcal{D}} d(O_i, \rho)$ up to precision $\epsilon/2$.

Theorem 3. Under the same assumptions as in Theorem 7 in Appendix C 2, except that we do not restrict the form of the initial state distribution $\mathcal{D}$, the hypothesis testing problem in Eq. (8) can be solved using $\widetilde{\mathcal{O}}(\Gamma T \epsilon^{-4} \log(\delta^{-1}) \log(\chi \delta^{-1}) \max_i \|O_i\|^2 \max_i \|O_i\|_{\text{shadow}}^2)$ samples.

So far, we have considered quantities that are on average conserved for an ensemble of states. It is natural to ask whether we can determine if an observable O is conserved

for all states, namely, $[H, O] = 0$. Through a quantum query complexity lower bound, we can show that this task cannot be accomplished efficiently in the worst case. The high-level idea of this argument goes as follows. Suppose that we have a black-box oracle U encoding a bit string $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$ through $U|n\rangle = (-1)^{\mathbf{x}_n}|n\rangle$; then letting $H = U$ we can implement e^{-iHt} using two queries to U. If an algorithm can distinguish between $\|[H, O]\| = 0$ or ≥ 1 with high probability with Q queries to e^{-iHt}, we can then show that it can evaluate OR($\mathbf{x}$) with $2Q$ queries to U. The query complexity lower bound of the OR function [61] then tells us that $Q = \Omega(2^{N/2})$. For a detailed statement of the result and its proof, see Appendix E.

IV. NUMERICAL EXPERIMENTS

In this section, we illustrate our algorithm with numerical examples. We consider a $\mathbb{Z}_2$ gauge theory and a disordered Heisenberg model in one dimension.

A. Identifying conservation laws in a lattice gauge theory

As a first numerical example, we consider a $\mathbb{Z}_2$ lattice gauge theory with staggered matter fields in one spatial dimension. For simplicity, we assume that the number of qubits N is even, $N \mod 2 = 0$. Then, the Hamiltonian is specified by

$$H_{\mathbb{Z}_2} = \frac{1}{2a} \sum_{i=0}^{N/2-1} (\sigma_{2i}^+ \sigma_{2i+1}^x \sigma_{2i+2}^- + \text{H.c.})$$
$$+ m \sum_{i=0}^{N/2-1} \frac{(-1)^i}{2} (\mathbb{I}_2 + \sigma_{2i}^z) + e \sum_{i=0}^{N/2-1} \sigma_{2i+1}^z, \quad (10)$$

where we choose periodic boundary conditions. By direct inspection of the Hamiltonian, we find that we can expect $N/2 + 2$ conservation laws, given by the Hamiltonian itself, the magnetization

$$M = \sum_{i=0}^{N/2-1} \sigma_{2i}^z, \quad (11)$$

and $N/2$ Gauss laws

$$G_{2j} = \sigma_{2j-1}^z \sigma_{2j}^z \sigma_{2j+1}^z. \quad (12)$$

We note that the magnetization and Hamiltonian are linear combinations of geometrically 1-local and 3-local terms, which have support on the entire system. In contrast, the $N/2$ Gauss laws are strictly geometrically 3-local. In the following, we consider $N = 8$ qubits, set the mass parameter to unity $m = 1$, and choose the electric field $e = 3/2m$ and lattice spacing $a = m/3$.
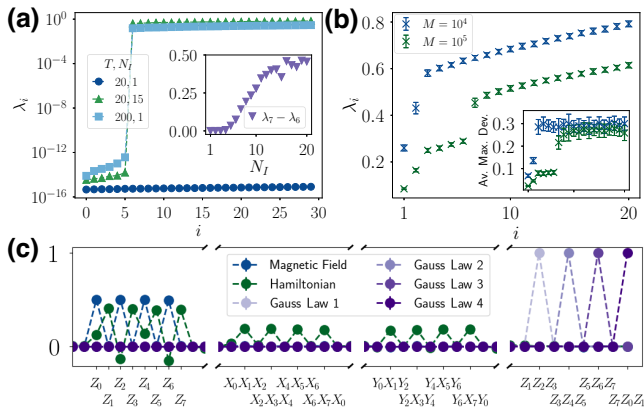
FIG. 1. Learning conservation laws in a $\mathbb{Z}_2$ gauge theory. In panel (a), we display the 30 smallest eigenvalues of the exact data matrix for various evolution times T and numbers of initial states N_I. At sufficiently long times T, $N + 2$ singular values are gapped out (light and dark blue). The use of multiple initial states (green triangles) allows us to shorten the evolution time considerably. Inset illustrates this effect, showing the gap as a function of the number of initial states for fixed time $T = 20$. In panel (b), we display the 20 smallest eigenvalues of the noisy data matrix, constructed from $M = 10^4$ and $M = 10^5$ per initial state ($N_I = 15$) and final time $T = 20$, $N_T = 41$. For $M = 10^4$ ($M = 10^5$), increasing M, two (six) singular values are gapped out. Testing these against independently obtained data yields small variations over time (inset). In panel (c), we display the Pauli basis expansion of the corresponding lowest six singular vectors. We identify the magnetization, Hamiltonian, and four Gauss laws. In all panels, $N = 8$ and $N_T = 2N_P/N_I = 624$. Points with error bars are the average and standard deviation over 25 experiments with identical parameters.

First, we investigate our protocol in the absence of shot noise due to a finite number of measurements [Fig. 1(a)]. We aim to learn all conservation laws with weight $k \leq 3$. We find that, for sufficiently long times $T = 200$, data collected from dynamics starting from a single initial random product state $N_I = 1$ is sufficient to identify all expected $N/2 + 2$ conservation laws: the singular values λ_i for $1 \leq i \leq N/2 + 2$ are close to zero (a nonzero value originates from finite machine precision) with a large gap to $\lambda_{N/2+3}$. At shorter times $T = 20$ and $N_I = 1$, the spectrum of singular values appears to be continuous, and conservation laws are not apparent. In contrast, data collected from several random initial product states is substantially more expressible (see Ref. [38] for a similar observation in the context of Hamiltonian learning). Even at short times $T = 20$, the expected conservation laws can be identified. We emphasize that, hereby, we keep the total number of points $N_T N_I \approx 2N_P$ where data are taken to be constant to enable a fair comparison ($N_T = 2N_P = 624$ for $N_I = 1$ and $N_T = 41$ for $N_I = 15$). In all cases, the time points are equally spaced.

Secondly, we simulate our complete protocol, including a finite number of measurements M per time point

and initial state. We choose $N_I = 15$ randomly chosen initial product states and a total evolution time of $T = 20$ with $N_T = 41$ steps. We find that, for a moderate number of $M = 10^5$ randomized measurements, $N/2 + 2$ singular values are gapped out [Fig. 1(b)]. Analyzing a Pauli basis expansion of the corresponding normalized singular vectors of our data matrix, we find that these correspond to the expected conservation laws, magnetization, energy (Hamiltonian), and $N/2$ Gauss laws [Fig. 1(c)].

Finally, to test that the learned quantities are indeed conserved over times, we employ an independent data set of the same size. We estimate expectation values of the learned quantities at different times and compute their maximum deviation from their mean values, averaged over initial states. We note that this represents a more simplified testing procedure than employed in Secs. II and III, and devote the full numerical implementation of the robust polynomial interpolation for testing to future work. Indeed, we find that quantities corresponding to small singular values have small variations over time up to shot noise originating from a finite number of randomized measurements M. The testing procedure also helps us better distinguish conservation laws from observables that are close to being conserved, by opening up the gap between singular values, as can be seen in Fig. 1(b). By using a different set of data to perform testing, we can exclude quantities with only small variation for a single noise realization or a single set of initial states. This is similar to detecting overfitting in supervised learning.

While we have so far concentrated on unitary dynamics, we emphasize that our protocol can serve to learn conservation laws of arbitrary quantum dynamics. To illustrate this point, we add local dephasing with strength γ, corresponding to jump operators $L_i = \sqrt{\gamma}\sigma_i^z$, and solve the corresponding Lindblad equation (all other Hamiltonian parameters remain the same). Since magnetization and Gauss laws are diagonal in the computational Z basis, they also remain conserved for $\gamma > 0$. In contrast, energy conservation is lost, as shown in Fig. 2(a) where λ_2 is missing (the indices of the subsequent singular values $\lambda_3, \lambda_4, \ldots$ have been shifted by 1 for clarity). We note that this shows that our protocol can serve as an indicator of decoherence mechanisms in an experiment targeting to implement specific (e.g., Hamiltonian) dynamics—by comparing learned versus expected conserved quantities. In this context, we remark that experimental imperfection in the implementation of the randomized measurement itself can, in principle, be detrimental to our protocol. However, powerful error mitigation techniques in terms of "robust classical shadows" exist that allow us to fully mitigate such measurement errors under very mild assumptions and with a small overhead in terms of the required number of measurements [55–57].

Our numerical experiments demonstrate that we can learn conservation laws, which are linear combinations
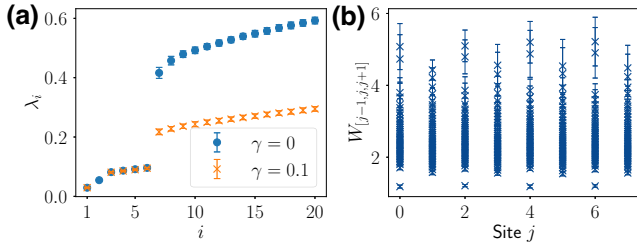
FIG. 2. Learning conservation laws in open quantum systems. In panel (a), we display the 20 smallest singular values of W for both the Hamiltonian (blue) and Lindbladian dynamics (orange) with local dephasing with rate $\gamma = 0.1$. While magnetization and Gauss laws are conserved in both cases, the Hamiltonian (energy) is only conserved for Hamiltonian dynamics (λ_2 is absent for $\gamma = 0.1$). We choose $M = 10^6$ measurements and use a Gaussian noise approximation to simulate the resulting shot noise. In all panels, $N = 8$, $N_T = 41$, $N_I = 15$, and points with error bars are the average and standard deviation over 25 experiments with identical parameters. Panel (b) displays the singular values obtained from the data matrix $W_{[j-1,j,j-1]}$ restricted to three adjacent subsystem sites as a function of j. This allows us to learn the Gauss laws, corresponding to the gapped singular values at even j, with considerably fewer measurements $M = 10^4$ [cf. Fig. 1(b)].

of k-local Pauli strings with a moderate number of randomized measurements M. We can decrease the required number of measurements further if we restrict ourselves to learning conservation laws with support on subsystems only, i.e., disregard quantities such as the magnetization or Hamiltonian, which are linear combinations of few-body terms, but have support on the entire system. To achieve this, we construct reduced data matrices W_A from measurement data obtained from subsystem A only. This is illustrated in Fig. 2(b), where we plot the singular values of data matrices W_{A_j} with $A_j = [j-1, j, j+1]$ containing the three sites $j-1, j, j+1$ as a function of j (periodic boundary conditions are implied). With only $M = 10^4$ randomized measurements [cf. $M = 10^5$ in Fig. 1(b)] per time point and initial state, we can identify the expected $N/2$ Gauss laws contained in subsystems A_j with $j \mod 2 = 0$.

B. Identifying conservation laws in a many-body-localized system

Next, we consider a disordered one-dimensional *XXZ*-spin chain with nearest-neighbor interactions that serves as a standard model for investigating many-body localization and thermalization [26–28]. It is described by the Hamiltonian

$$H_{XXZ} = J_x \sum_i (\sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y) + J_z \sum_i \sigma_i^z \sigma_{i+1}^z$$
$$+ \sum_i h_i \sigma_i^z, \quad (13)$$

where the local disorder potentials h_i $(i = 1, \ldots, N)$ are randomly distributed in the interval $[-w, w]$, and w is the disorder strength. We assume in the following that $J_x = J_z = 1$.

Numerical studies [18,62–64] indicate that this model exhibits a transition from an ergodic phase at weak disorder to an ergodicity breaking many-body-localized phase at sufficiently strong disorder [65]. While in the ergodic phase, nearly all eigenstates obey the eigenstate thermalization hypothesis [66,67], in the MBL phase the eigenstate thermalization hypothesis is not valid, and the system is characterized by an extensive number of *quasilocal* conservation laws τ_i $(i = 1, \ldots, N)$, called l bits [18–21]. In terms of these conservation laws, Hamiltonian (13) can be written as

$$H_{\text{diag}} = \sum_i \xi_i \tau_i + \sum_{i<j} J_{ij} \tau_i \tau_j + \sum_{i<j<k} J_{ikj} \tau_i \tau_j \tau_k + \cdots,$$
$$(14)$$

where, in the MBL phase, the τ_i are quasilocal, in the sense that they can be approximated by geometrically local operators to exponential precision, and the coupling coefficients J_{ij}, J_{ikj} decay exponentially with distance $|i - j|$. We emphasize that, by switching to an energy eigenbasis, we can always rewrite Hamiltonian (13) in the form of Eq. (14), also in the thermalizing phase. In general, however, the conservation laws τ_i will be completely nonlocal, high-weight operators with vanishing overlap to the microscopic degrees of freedom σ_i^z. In this case, Eq. (14) is of little use. Finally, we note that, independent of the disorder strength, there are always two conserved quantities given as sums of local observables, which are the Hamiltonian itself and the total magnetization.

With our learning algorithm, we target conserved quantities given as sums of local observables. While we expect to be able to learn the local *XXZ* Hamiltonian and total magnetization for all disorder strengths, the conserved quantities τ_i are expected to be inaccessible in the thermal phase due to their nonlocal nature. In contrast, for strong disorder, we expect an extensive amount of approximately conserved local quantities, approximating the quasilocal l bits τ_i to high precision. To test these expectations in small systems, we simulate our protocol by sampling the random initial product state, picking a random disorder pattern, and time evolving under the corresponding *XXZ* Hamiltonian (13) using exact diagonalization. We evaluate Pauli expectations of all N_P geometrically up to 3-local Pauli operators at $N_T = 41$ equidistant time points up to a final time $T = 40$. We repeat this for $N_I \approx 2N_P/N_T$ initial states per fixed disorder pattern. To simulate realistic shot noise arising from a finite number of randomized measurements, we use a Gaussian noise approximation adding independent Gaussian noise to each expectation value with a variance corresponding to $M = 5 \times 10^5$ randomized measurements
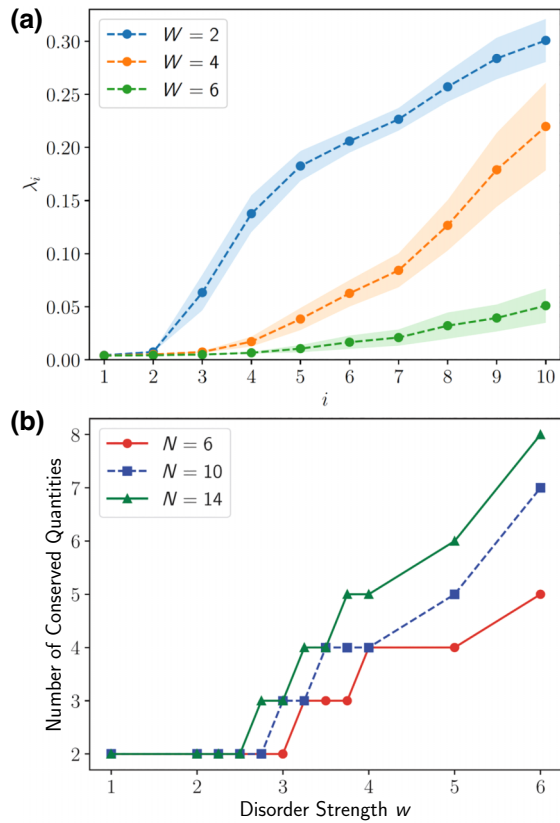
FIG. 3. Learning conservation laws in disordered spin chains. (a) We display the ten lowest squared singular values of the data matrix $\hat{W}$ constructed using $N_I = 17$ initial states evolved according to H_{XXZ} with different disorder strengths $w = 2, 4, 6$ (blue, orange, green) to a final time $T = 40$ and evaluated at $N_T = 41$ equally spaced time points. We add Gaussian noise, simulating shot noise arising from $M = 5 \times 10^5$ measurements. Each point is averaged over 11 random disorder patterns; the error bars indicate the standard error of the mean. The two smallest singular values λ_1, λ_2 correspond to the magnetization and Hamiltonian, respectively. As the disorder strength w becomes larger, the singular values $\lambda_{i \geq 3}$ decrease significantly, indicating an increasing number of approximate conservation laws. (b) We show the number of singular values below a threshold $\epsilon = 0.02$ as a function of w for different system sizes. As in panel (a), we choose $M = 5 \times 10^5, T = 40, N_T = 41$. The numbers of initial states for system sizes $N = 6, 10, 14$ are $N_I = 9, 17, 24$, respectively, to ensure that $N_I N_T \approx 2 N_P$. Points correspond to the median of 11 random Hamiltonian configurations. We observe a sharp increase with increasing disorder strength.

per time point and initial state [68]. Finally, we construct the data matrix $\hat{W}$ and perform SVD.

In Fig. 3(a), we display the ten lowest singular values of $\hat{W}$ as a function of disorder strength, with each point corresponding to an average over 11 random Hamiltonian configurations. Consistent with our expectation, we find, independent of the disorder strength, two small singular values, corresponding to the magnetization and Hamiltonian. As remarked before, they still attain a nonzero value

due to the finite number of measurements M. In addition, we find that the singular values λ_i ($i \geq 3$) decrease strongly with increasing disorder strength, indicating an increasing number of approximately conserved quantities. To show this more quantitatively, in Fig. 3(b) we plot the number of singular values below a threshold $\epsilon = 0.02$ as a function of the disorder strength for various system sizes. We observe a sharp increase at a disorder strength, which is consistent with previous findings on the onset of many-body localization effects in finite-size systems [18,62–64]. In addition, the number of singular values below the threshold increases with system size, indeed indicating an extensive number of approximately conserved quantities.

V. CONCLUSION AND OUTLOOK

In this paper, we propose a method for learning conservation laws in arbitrary quantum dynamics. Our method can find a set of observables that include all conservation quantities with high probability, and we also propose a method to test the conservation law candidates obtained in this way. The sample complexity and classical processing time are both at most polynomial in the system size. The conservation laws hold for either a single input state or an ensemble of input states, and for the latter case, we derive a generalization bound ensuring that the result from finitely many samples can be reliably generalized to the entire ensemble. We provided a proof of principle of our method using numerical experiments in a one-dimensional $\mathbb{Z}_2$ lattice gauge theory and one-dimensional MBL systems. Beyond these examples, we envision a wide range of applications for our protocol, ranging from Hilbert space fragmentation [29] to random circuits with symmetries [69] and the study of general quantum channels [70]. In addition, knowledge of conserved quantities in dynamics enables powerful error mitigation techniques for noisy intermediate-scale quantum devices [71] and more efficient (randomized measurement) protocols for probing many-body entanglement [72,73].

ACKNOWLEDGMENTS

APPENDIX A: COUNTING CONSERVATION LAWS

In this appendix, we prove that the learning procedure outlined in Sec. II can reliably provide an upper bound for the number of conserved quantities. To simplify the notation, we consider mostly the case of a single initial state, but comment on the appropriate redefinitions for the case of multiple initial states. The proof generalizes to the latter situation without requiring any change.

We define the expectation value matrix $X = (X_{ij})_{N_P \times N_T}$ (we require that $N_T \geq N_P$), where

$$X_{ij} = \langle P_i(t_j) \rangle. \tag{A1}$$

From experiments we obtain estimates forming a matrix $\hat{X} = (\hat{X}_{ij})_{N_P \times N_T}$. The error is described by the matrix

$$E_{ij} = \hat{X}_{ij} - X_{ij}. \tag{A2}$$

The estimate is unbiased, which means that $\mathbb{E}[E] = 0$. The variance depends on what method we use to get the estimates $\hat{X}_{ij}$. In this appendix, we discuss two scenarios: the naive approach, which requires repreparing the state for each expectation value, and classical shadows, enabling the simultaneous estimation of many expectation values.

Matrix X can be used to characterize conservation laws. If an observable

$$\sum_i c_i P_i \tag{A3}$$

is conserved then we have

$$\sum_i c_i \langle P_i(t_j) \rangle = \frac{1}{N_T} \sum_{ij'} c_i \langle P_i(t'_j) \rangle \tag{A4}$$

for all $j = 1, 2, \ldots, N_T$. Writing the above equation in matrix form, we have

$$c^\top X = \frac{1}{N_T} c^\top X \mathbf{1} \mathbf{1}^\top, \tag{A5}$$

where $\mathbf{1} = (1, 1, \ldots, 1)^\top$ is the N_T-dimensional vector with all entries being 1. Therefore, every conserved quantity is contained in the null space of the matrix

$$W^\top = \left(I - \frac{1}{N_T} \mathbf{1} \mathbf{1}^\top\right) X^\top, \tag{A6}$$

where I denotes the identity matrix. The dimension of its null space, which we denote by D_{null}, provides an upper bound for the number of conserved quantities.

When we use multiple initial states, the relationship between W and X becomes slightly different. The entry-wise relationship is $W_{i,jk} = X_{i,jk} - N_T^{-1} \sum_{j'} X_{i,j'k}$, which we can reformulate in matrix language as

$$W^\top = \left[\left(I - \frac{1}{N_T} \mathbf{1} \mathbf{1}^\top\right) \otimes I\right] X^\top, \tag{A7}$$

where $I - \mathbf{1} \mathbf{1}^\top / N_T$ acts on the index j and I acts on the index k.

Note that D_{null} is not directly available to us. What we can do is estimate D_{null} through the number of small singular values of the matrix

$$\hat{W}^\top = \left(I - \frac{1}{N_T} \mathbf{1} \mathbf{1}^\top\right) \hat{X}^\top. \tag{A8}$$

When there are multiple initial states, the transformation from $\hat{X}$ to $\hat{W}$ is similar to Eq. (A7). Note that in both cases, because $\|I - \mathbf{1} \mathbf{1}^\top / N_T\| \leq 1$, we have $\|W - \hat{W}\| \leq \|X - \hat{X}\| = \|E\|$. We denote the number of singular values of the above matrix that are below ϵ by $\hat{D}_{\text{null}}$. We show that, with enough samples, we can guarantee that, with large probability,

$$D_{\text{null}} \leq \hat{D}_{\text{null}}. \tag{A9}$$

We denote the singular values of W and $\hat{W}$ by σ_i and $\hat{\sigma}_i, j = 1, 2, \ldots, N_P$, arranged in ascending order, respectively. By Mirsky's inequality [74], we have

$$|\hat{\sigma}_i - \sigma_i| \leq \|W - \hat{W}\| \leq \|E\|, \tag{A10}$$

where $\|\cdot\|$ denotes the spectral norm. Consequently, if $\sigma_i = 0$ for any i then $\hat{\sigma}_i \leq \|E\|$. Therefore, we have the following lemma.

Lemma 1. When $\|E\| \leq \epsilon$, inequality (A9) holds.

Furthermore, we do not need the upper bound for $\|E\|$ to hold with probability 1. Rather, as long as the upper bound holds with a probability that is greater than $1/2$ by a constant, we can simply repeat the procedure multiple times, and take the median of all $\hat{D}_{\text{null}}$ that are computed. This ensures that the median $\hat{D}_{\text{null}}^{\text{median}} \geq D_{\text{null}}$ with probability at least $1 - \delta$ with $\mathcal{O}(\log(\delta^{-1}))$ using the Chernoff bound. Consequently, it suffices to upper bound $\mathbb{E}\|E\|$.

Lemma 2. When $\mathbb{E}\|E\| \leq \epsilon/4$, then

$$D_{\text{null}} \leq \hat{D}_{\text{null}}^{\text{median}} \tag{A11}$$

with probability at least $1 - \delta$, where $\hat{D}_{\text{null}}^{\text{median}}$ is the median taken over $\mathcal{O}(\log(\delta^{-1}))$ independent samples of $\hat{D}_{\text{null}}$.

Our main tool in bounding $\mathbb{E}\|E\|$ is through the non-commutative Khintchine inequality [75, Section 9.8]. This inequality implies that, as stated in Ref. [76, Eq. (1.2)], for a real symmetric random matrix $M = \sum_j g_j A_j$, where the g_j are independent and identically distributed standard Gaussian random variables and the A_j are real symmetric matrices of size $d \times d$, we have

$$C_1 \sqrt{\left\| \sum_j A_j^2 \right\|} \leq \mathbb{E}\|M\| \leq C_2 \sqrt{\left\| \sum_j A_j^2 \right\| \log(d)}. \tag{A12}$$

Because of the requirement for M to be real and symmetric, instead of directly considering E, we need to consider

$$S_E = \begin{pmatrix} 0 & E \\ E^\top & 0 \end{pmatrix} = \sigma^- \otimes E + \sigma^+ \otimes E^\top. \tag{A13}$$

Note that $\|S_E\| = \|E\|$. Therefore, we only need to upper bound $\|S_E\|$.

1. The naive approach

Let us first consider the approach where each entry of $\hat{X}$ is sampled independently. In this scenario, the entries of matrix E are independent. Therefore, we can write

$$E = \sum_{ij} g_{ij} \sigma_{ij} e_i e_j^\top. \tag{A14}$$

Here, g_{ij} is a standard Gaussian random variable and

$$\sigma_{ij}^2 = \frac{\langle P_i(t_j)^2 \rangle - \langle P_i(t_j) \rangle^2}{N_s}, \tag{A15}$$

where N_s is the number of samples used for each $\hat{X}_{ij}$. We are assuming that the error is Gaussian, which is reasonable for large N_s due to the central limit theorem.

Correspondingly,

$$S_E = \sum_{ij} g_{ij} \sigma_{ij} (\sigma^- \otimes e_i e_j^\top + \sigma^+ \otimes e_j e_i^\top). \tag{A16}$$

By Eq. (A12) we have

$$\mathbb{E}[\|S_E\|] \leq C_2 \sqrt{\left\| \sum_{ij} \sigma_{ij}^2 (|0\rangle\langle 0| \otimes e_i e_i^\top + |1\rangle\langle 1| \otimes e_j e_j^\top) \right\| \log(N_P + N_T)}$$

$$\leq C_2 \sqrt{\max\left\{ \max_i \sum_j \sigma_{ij}^2, \max_j \sum_i \sigma_{ij}^2 \right\} \log(N_P + N_T)}. \tag{A17}$$

Because $\|P_i\| \leq 1$, we have

$$\mathbb{E}[\|E\|] \leq \mathbb{E}[\|S_E\|] \leq \mathcal{O}\left(\sqrt{\frac{N_P + N_T}{N_s} \log(N_P + N_T)} \right). \tag{A18}$$

To ensure that $\mathbb{E}[\|E\|] \leq \epsilon/4$, we need to choose

$$N_s = \widetilde{\mathcal{O}}((N_P + N_T)\epsilon^{-2}). \tag{A19}$$

The total number of samples is therefore

$$N_P \times N_T \times N_s \times \mathcal{O}(\log(\delta^{-1}))$$
$$= \widetilde{\mathcal{O}}(N_P N_T (N_P + N_T)\epsilon^{-2} \log(\delta^{-1})), \tag{A20}$$

where we recall that N_P is the number of geometrically k-local Pauli observables, N_T is the number of time points, and N_s is the number of samples for estimating each matrix element $\hat{X}_{ij}$.

When considering N_I initial states, we replace all N_T with $N_T N_I$. The above expression then becomes

$$N_P \times N_T N_I \times N_s \times \mathcal{O}(\log(\delta^{-1}))$$
$$= \widetilde{\mathcal{O}}(N_P N_T N_I (N_P + N_T N_I)\epsilon^{-2}\log(\delta^{-1})). \quad \text{(A21)}$$

Through Lemma 2, we can compute the cost of learning.

Theorem 4. With $\widetilde{\mathcal{O}}(N_P N_T N_I (N_P + N_T N_I)\epsilon^{-2}\log(\delta^{-1}))$ samples, we can ensure that $D_{\text{null}} \leq \hat{D}_{\text{null}}^{\text{median}}$ with probability at least $1 - \delta$, where $\hat{D}_{\text{null}}^{\text{median}}$ is the median taken over $\mathcal{O}(\log(\delta^{-1}))$ independent samples of $\hat{D}_{\text{null}}$.

$$\Sigma_{ii'}^{j} = \begin{cases} 3^{\omega(P_i,P_{i'})}\langle P_i(t_j)P_{i'}(t_j)\rangle - \langle P_i(t_j)\rangle\langle P_{i'}(t_j)\rangle & \text{if } P_i \text{ and } P_{i'} \text{ completely commute,} \\ -\langle P_i(t_j)\rangle\langle P_{i'}(t_j)\rangle & \text{otherwise,} \end{cases} \quad \text{(A22)}$$

where by "completely commute" we mean that the two Pauli operators can be simultaneously diagonalized in the same single-qubit Pauli eigenbasis, and $\omega(P_i, P_{i'})$ is the number of qubits on which P_i and $P_{i'}$ overlap. Let us first perform an eigendecomposition for Σ^j:

$$\Sigma^j = \sum_l \lambda_l^j v_l^j v_l^{j\top} \quad \text{(A23)}$$

with $\lambda_l^j \geq 0$ because Σ^j is symmetric positive semidefinite. Then we have

$$\hat{X}_{\cdot j} = \sum_l g_{lj}\sqrt{\lambda_l^j/N_s}\, v_l^j + X_{\cdot j}, \quad \text{(A24)}$$

where "=" means equal in distribution, and the g_{lj} are independent and identically distributed standard Gaussian random variables. Then, the error matrix E can be written as

$$E = \sum_{lj} g_{lj}\sqrt{\lambda_l^j/N_s}\, v_l^j e_j^\top. \quad \text{(A25)}$$

Through the same analysis as in Eq. (A17), we have

$$\mathbb{E}[\|S_E\|]$$
$$\leq C_2\sqrt{\max\left\{\max_i \sum_j \frac{\lambda_l^j}{N_s}, \max_j \sum_i \frac{\lambda_l^j}{N_s}\right\}\log(N_P + N_T)}$$
$$\leq C_2\sqrt{\max_j \|\Sigma^j\|\frac{N_P + N_T}{N_s}\log(N_P + N_T)}. \quad \text{(A26)}$$

Note that, in practice, we usually do not choose N_I to be large, but rather choose $N_T N_I = \mathcal{O}(N)$, where N is the system size.

2. Using classical shadows

Using classical shadows to construct $\hat{X}$, we no longer have the simple decomposition in Eq. (A14). At each time t_j, the vector consisting of observable expectation values $\hat{X}_{\cdot j}$ is a Gaussian random vector with covariance matrix Σ^j/N_s (again, Gaussianity is a result of the central limit theorem), in which

Therefore,

$$\mathbb{E}[\|E\|] \leq \mathcal{O}\left(\sqrt{\max_j \|\Sigma^j\|\frac{N_P + N_T}{N_s}\log(N_P + N_T)}\right). \quad \text{(A27)}$$

The next step is to then bound $\max_j \|\Sigma^j\|$. Note that in the worst case, we have $\max_j \|\Sigma^j\| = \mathcal{O}(N_P)$. This is in fact attainable: we can choose $\rho(t_j) = |\text{GHZ}\rangle\langle\text{GHZ}|$, where $|\text{GHZ}\rangle = (|00\cdots0\rangle + |11\cdots1\rangle)/\sqrt{2}$, and let $P_i = Z_i$ for $i = 1, 2, \ldots, N_P$. Then, we have $\Sigma^j = 2I + \mathbf{1}\mathbf{1}^\top$, thus giving $\|\Sigma^j\| = \mathcal{O}(N_P)$.

In this worst case, in order to ensure that $\mathbb{E}[\|E\|] \leq \epsilon/4$, we need

$$N_s = \mathcal{O}(N_P(N_P + N_T)\log(N_P + N_T)\epsilon^{-2}). \quad \text{(A28)}$$

The total number of samples needed is

$$N_T \times N_s \times \mathcal{O}(\log(\delta^{-1}))$$
$$= \widetilde{\mathcal{O}}(N_P N_T (N_P + N_T)\epsilon^{-2}\log(\delta^{-1})). \quad \text{(A29)}$$

Note that here, even though we did not need to multiply by N_P on the left-hand side as in Eq. (A20), an N_P dependence nevertheless appears on the right-hand side through N_s. As a result, we still get the same sample complexity scaling as in the naive approach [Eq. (A20)].

However, if correlation decays rapidly, classical shadows can offer us an advantage. More specifically, let us assume that the quantum system is on a D-dimensional

lattice. Furthermore, for all $0 \leq t \leq T$,

$$| \langle P_i(t)P_{i'}(t)\rangle - \langle P_i(t)\rangle \langle P_{i'}(t)\rangle | \leq \mathcal{O}(e^{-d(P_i,P_{i'})/\xi}), \quad \text{(A30)}$$

where $d(P_i, P_{i'})$ is the distance between P_i and $P_{i'}$, and ξ is the correlation length. Because the P_i are supported on at most $k = \mathcal{O}(1)$ adjacent qubits, the number of $P_{i'}$ within r distance from P_i grows like r^D. Therefore we have

$$\sum_{i'} | \langle P_i(t)P_{i'}(t)\rangle - \langle P_i(t)\rangle \langle P_{i'}(t)\rangle | \leq \mathcal{O}(\xi^D) \quad \text{(A31)}$$

for all i. For $\Sigma_{ii'}^j$, in each row there are only $\mathcal{O}(1)$ many entries that are different from $\langle P_i(t)P_{i'}(t)\rangle - \langle P_i(t)\rangle \langle P_{i'}(t)\rangle$, as can be seen from Eq. (A22). The absolute value of each entry is upper bounded by $\mathcal{O}(1)$. Consequently,

$$\sum_{i'} \|\Sigma_{ii'}^j\| \leq \mathcal{O}(\xi^D). \quad \text{(A32)}$$

Then we have

$$\|\Sigma^j\| \leq \max_i \sum_{i'} \|\Sigma_{ii'}^j\| \leq \mathcal{O}(\xi^D), \quad \text{(A33)}$$

which indicates that $\|\Sigma^j\| = \mathcal{O}(1)$ when $\xi, D = \mathcal{O}(1)$.

In this good scenario, we only need

$$N_s = \mathcal{O}((N_P + N_T) \log(N_P + N_T)\epsilon^{-2}). \quad \text{(A34)}$$

The total number of samples needed is

$$N_T \times N_s \times \mathcal{O}(\log(\delta^{-1})) = \widetilde{\mathcal{O}}(N_T(N_P + N_T)\epsilon^{-2} \log(\delta^{-1})), \quad \text{(A35)}$$

which is quadratically better than the scaling in terms of the N_P dependence. When we take into account having N_I initial states, the number of samples then becomes

$$N_T N_I \times N_s \times \mathcal{O}(\log(\delta^{-1}))$$
$$= \widetilde{\mathcal{O}}(N_T N_I(N_P + N_T N_I)\epsilon^{-2} \log(\delta^{-1})). \quad \text{(A36)}$$

Again, through Lemma 2, we have the following result.

Theorem 5. We assume that the quantum system is defined on a D-dimensional lattice and that Eq. (A30) holds for all $0 \leq t \leq T$. Then, with $\widetilde{\mathcal{O}}(N_T N_I(N_P + N_T N_I)\epsilon^{-2} \log(\delta^{-1}))$ samples, we can ensure that $D_{\text{null}} \leq \hat{D}_{\text{null}}^{\text{median}}$ with probability at least $1 - \delta$, where $\hat{D}_{\text{null}}^{\text{median}}$ is the median taken over $\mathcal{O}(\log(\delta^{-1}))$ independent samples of $\hat{D}_{\text{null}}$.

APPENDIX B: ACCURACY OF THE LEARNED CONSERVATION LAWS

In the learning procedure described in Sec. II, we obtain a subspace spanned by the singular vectors of matrix $\hat{W}$. We show in this appendix that this subspace contains all the conserved quantities approximately.

We use the following result for singular vector perturbation to characterize the accuracy of the conservation laws obtained from our learning procedure.

Lemma 3. Suppose that we have matrix $A \in \mathbb{R}^{M \times N}$ and vector $w \in \mathbb{R}^M$ such that $w^\top A = 0$. We let $\hat{A} = A + \delta A$. We then write down the singular value decomposition of $\hat{A}$ as $\hat{A} = \sum_{k=1}^r \hat{\sigma}_k \hat{u}_k \hat{v}_k^\top$, where $\hat{\sigma}_1 \leq \cdots \leq \hat{\sigma}_r$. Let r' be the largest index such that $\hat{\sigma}_{r'} \leq \epsilon$. We decompose w through

$$w = \hat{w} + w_\perp, \quad \text{(B1)}$$

where $\hat{w} = \sum_{k=1}^{r'} \hat{u}_k \hat{u}_k^\top w$. Then we have

$$\|w_\perp\| \leq \frac{\|w^\top \delta A\|}{\epsilon}. \quad \text{(B2)}$$

Proof. By the triangle inequality

$$\|w^\top \hat{A}\| \leq \|w^\top A\| + \|w^\top \delta A\| = \|w^\top \delta A\|. \quad \text{(B3)}$$

On the other hand,

$$\|w^\top \hat{A}\|^2 = \hat{w}^\top \hat{A}\hat{A}^\top \hat{w} + \hat{w}^\top \hat{A}\hat{A}^\top w_\perp + w_\perp^\top \hat{A}\hat{A}^\top \hat{w}$$
$$+ w_\perp^\top \hat{A}\hat{A}^\top w_\perp$$
$$\geq w_\perp^\top \hat{A}\hat{A}^\top w_\perp, \quad \text{(B4)}$$

where we have used the facts that $\hat{w}^\top \hat{A}\hat{A}^\top \hat{w} \geq 0$ and

$$\hat{w}^\top \hat{A}\hat{A}^\top w_\perp = \sum_{k=1}^r \hat{\sigma}_k^2 \hat{w}^\top \hat{u}_k \hat{u}_k^\top w_\perp = 0 \quad \text{(B5)}$$

because of the orthogonal decomposition (B1). Similarly, we have $w_\perp^\top \hat{A}\hat{A}^\top \hat{w} = 0$. Because $w_\perp$ only overlaps with $\hat{u}_k$ for $k \geq r' + 1$, we have

$$w_\perp^\top \hat{A}\hat{A}^\top w_\perp \geq \hat{\sigma}_{r'+1}^2 \|w_\perp\|^2 \geq \epsilon^2 \|w_\perp\|^2. \quad \text{(B6)}$$

Therefore,

$$\|w^\top \hat{A}\|^2 \geq \epsilon^2 \|w_\perp\|^2. \quad \text{(B7)}$$

Combining the above with Eq. (B3) we have

$$\|w_\perp\| \leq \frac{\|w^\top \delta A\|}{\epsilon}. \quad \text{(B8)}$$

This completes the proof. ∎

In the context of our algorithm, we let $A = W$, $\hat{A} = \hat{W}$, and $w = \vec{c}$ ($\|\vec{c}\| = 1$) in the above lemma. Here $\vec{c}$ corresponds to an exact conserved quantity $O = \sum_i c_i P_i$, $\hat{W}$ is the shifted data matrix, in which we subtract the time average from all entries so that each row sums to zero, and W is its noiseless limit. For simplicity, we first focus on the case with a single initial state, i.e., $N_I = 1$. In practice, W is perturbed to be $\hat{W}$, and the corresponding perturbation is $\delta A = E(I - \mathbf{1}\mathbf{1}^\top/N_T)$, where E contains the noise on each entry of the data matrix X. The above result tells us that the subspace we obtain through performing SVD on $\hat{W}$ approximately contains the exact conservation law O if

$$\|\vec{c}^\top E(I - \mathbf{1}\mathbf{1}^\top/N_T)\| \ll \epsilon. \tag{B9}$$

Here, ϵ is our chosen truncation threshold for singular values. To see what this means for the conservation law $O = \sum_i c_i P_i$, we first define

$$\epsilon' = \|\vec{c}^\top E(I - \mathbf{1}\mathbf{1}^\top/N_T)\|/\epsilon \ll 1.$$

Then, by Lemma 3, there exists $\hat{\vec{c}}$ within the subspace of singular vectors that we obtain from SVD such that

$$\|\vec{c} - \hat{\vec{c}}\| \le \epsilon'.$$

Note that $\hat{\vec{c}} = (\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_{N_P})$ corresponds to an operator $\hat{O} = \sum_i \hat{c}_i P_i$ that we learn. The above bound tells us that

$$\|\hat{O} - O\|_{\mathrm{HS}} \le \epsilon', \tag{B10}$$

where $\|\cdot\|_{\mathrm{HS}}$ denotes the normalized Hilbert-Schmidt norm, i.e., $\|M\|_{\mathrm{HS}} = \mathrm{Tr}[M^\dagger M]/2^N$. Therefore, we learn operator $\hat{O}$ from performing SVD is an approximate conservation law corresponding to O.

The above discussion is still valid for multiple initial states, i.e., $N_I > 1$. Instead of $\delta A = E(I - \mathbf{1}\mathbf{1}^\top/N_T)$, we have

$$\delta A = E[(I - \mathbf{1}\mathbf{1}^\top/N_T) \otimes I]$$

through Eq. (A7). All the subsequent derivation remains valid.

From the above analysis, we can see a tension in our choice of threshold ϵ: decreasing ϵ helps us better distinguish exactly conserved quantities from the approximate ones, but on the other hand, it increases the precision requirement on our data matrix $\hat{W}$.

APPENDIX C: TESTING CONSERVATION LAWS FOR A SINGLE INITIAL STATE

In this appendix, we discuss how to test the conserved quantities that we have learned. The testing procedure is briefly outlined in Sec. II. Here, we provide a more detailed description, prove its correctness, and also analyze the cost.

Let $f_i(t) = \mathrm{Tr}[\rho(t)O_i]$, where each O_i is a sum of low-weight Pauli operators, for $i = 1, 2, \ldots, \chi$. We further assume that, using the notation $f^{(k)}(t)$ to denote the kth derivative of $f(t)$,

$$|f_i^{(k)}(t)| \le \mathcal{C}\Gamma^k k! \tag{C1}$$

with constants $\mathcal{C}$ and Γ. We note that this is a very reasonable assumption to make. For time evolution under the von Neumann equation, this assumption holds with $\Gamma = \mathcal{O}(1)$ for geometrically local Hamiltonians and Hamiltonians with certain fast-decaying long-range interactions. For details, see Appendix F. In general, we can always choose $\Gamma = \|H\|$. For the Lindblad master equation, a similar result can also be obtained.

1. Expectation value interpolation for multiple observables

In this subsection, we find functions $\hat{p}_i(t)$, which are piecewise polynomials, such that

$$|f_i(t) - \hat{p}_i(t)| \le \epsilon \tag{C2}$$

with probability at least $1 - \delta$ for each $t \in [0, T]$.

a. Introduction to the robust polynomial interpolation algorithm

Here we provide a short high-level introduction to the robust polynomial interpolation algorithm used in our protocol. We use the result from Ref. [59] to quantify the number of samples needed to achieve a certain interpolation accuracy, but, in practice, we can use the conventional Chebyshev fitting method rather than the method proposed in Ref. [59]. This is because Kane *et al.* [59] carefully designed their algorithm to be robust against adversarial noise, whereas the noise we encounter from experiments is stochastic rather than adversarial.

In our protocol, the robust polynomial interpolation is only used for short-time interpolation. More precisely, we want to construct a polynomial $p(t)$ that approximates $f(t)$ uniformly in the interval $[0, T]$ for small T. With the assumption given in Eq. (C1), we consider the case where $\Gamma T \le 1$. For convenience, we define

$$\tilde{f}(x) = f((T/2)(x+1)), \tag{C3}$$

so that we have the variable x in the interval $[-1, 1]$, which is the domain on which Chebyshev polynomials form an orthogonal basis. We then expand $\tilde{f}(x)$ as a linear

combination of Chebyshev polynomials

$$\tilde{f}(x) = \sum_{k=0}^{\infty} c_k T_k(x), \tag{C4}$$

where $T_k(x)$ is the kth Chebyshev polynomial of the first kind. The coefficients can be obtained through

$$c_k = \frac{2 - \delta_{0k}}{\pi} \int_{-1}^{1} \frac{T_k(x)\tilde{f}(x)}{\sqrt{1-x^2}} \mathrm{d}x, \tag{C5}$$

where δ_{mn} denotes the Kronecker delta.

With the above setup, we can evaluate c_k numerically. First, we argue that one only needs to evaluate very few c_k. This is because we can truncate the Chebyshev expansion in Eq. (C4) at $K = \mathcal{O}(\log(\epsilon^{-1}))$ [as a result of Eq. (C1)] to get

$$\tilde{f}(x) = \sum_{k=0}^{K} c_k T_k(x) + \mathcal{O}(\epsilon). \tag{C6}$$

This can be proved by converting the Chebyshev expansion to the Fourier expansion, changing the variable $x = \cos(\theta)$, and then applying results about the decay of Fourier coefficients as discussed in Chapter 4 of Ref. [77]. We therefore only need to evaluate c_k for $k = 0, 1, \ldots, K$ to get accuracy $\mathcal{O}(\epsilon)$.

We then discuss two ways to evaluate c_k. The first is through Monte Carlo integration, and is the approach taken in Ref. [59]. Let $\hat{x}$ be a random variable obeying the Chebyshev distribution, i.e., with probability density $1/(\pi\sqrt{1-x^2})$ in the interval $[-1, 1]$. Then we can observe that

$$(2 - \delta_{0k})\mathbb{E}[T_k(\hat{x})\tilde{f}(\hat{x})] = c_k. \tag{C7}$$

Therefore we can generate multiple samples of $\hat{x}$, denoted by $x_1, x_2, \ldots, x_{N_s}$, and then approximate c_k by

$$c_k \approx \frac{2 - \delta_{0k}}{N_s} \sum_{j=1}^{N_s} T_k(x_j)\tilde{f}(x_j). \tag{C8}$$

Approximating each c_k to precision ϵ takes $N_s = \mathcal{O}(\epsilon^{-2})$ samples of $\hat{x}$ and the corresponding $\tilde{f}(\hat{x})$. When applying this approach to learn conservation laws, $f(t)$ and therefore $\tilde{f}(\hat{x})$ can only be evaluated with shot noise. This adds to the variance, but does not change the $\mathcal{O}(\epsilon^{-2})$ scaling. To ensure a uniform ϵ approximation for $f(t)$, we need each c_k to be evaluated to precision $\mathcal{O}(\epsilon/K)$. Note that $K = \mathcal{O}(\log(\epsilon^{-1}))$. We therefore arrive at a sample complexity of $\mathcal{O}(\epsilon^{-2}\log^2(\epsilon^{-1}))$ with this method; i.e., with this many pairs of $(t_j, f(t_j))$ where the value of $f(t_j)$
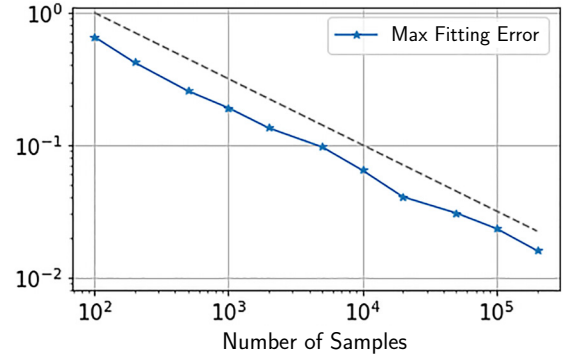

FIG. 4. The uniform approximation error $\max_{0 \le t \le T} |f(t) - p(t)|$ as a function of the number of samples of $f(t)$ using the Gauss-Chebyshev approach. Both axes are on the logarithmic scale. The gray dashed line has slope -0.5, representing the $1/\sqrt{N_s}$ scaling, where N_s is the number of samples. The function being fitted is $\tilde{f}(x) = (1 + i)e^{(0.3)i\pi x} + e^{-(0.19)i\pi x}$ with $K = 3$.

is subject to random noise, we will be able to obtain a polynomial $p(t)$ such that

$$\max_{t \in [0,T]} |f(t) - p(t)| \le \epsilon \tag{C9}$$

with large probability.

Another method for evaluating c_k is through the Gauss-Chebyshev quadrature. We can discretize the integral in Eq. (C5) as

$$c_k \approx \frac{2 - \delta_{0k}}{K+1} \sum_{j=1}^{K+1} T_k(x_j)\tilde{f}(x_j), \tag{C10}$$

where $x_j = \cos(2\pi(j-1)/(K+1))$. Note that because of the spectral accuracy of the Gauss-Chebyshev quadrature we only need $K + 1$ quadrature points instead of N_s different x values as in Eq. (C8). We therefore need to evaluate $\tilde{f}(x_j)$ to precision $\mathcal{O}(\epsilon/K)$ for each j. This takes $\mathcal{O}(K^2\epsilon^{-2})$ samples for each j. In total there are $K + 1$ different $\tilde{f}(x_j)$ to evaluate, and, consequently, the final sample complexity is $\mathcal{O}(K^3\epsilon^{-2}) = \mathcal{O}(\epsilon^{-2}\log^3(\epsilon^{-1}))$. Despite the slightly worse asymptotic scaling compared to the Monte Carlo integration approach, we found that if ϵ is not too small, the Gauss-Chebyshev method in fact takes fewer samples to achieve the same accuracy. In Fig. 4, we numerically show that the error decays roughly as $1/\sqrt{N_s}$ as the number of samples N_s increases.

b. Short-time interpolation

We first propose a method for the case where $\Gamma T \le 1$. In this case, $f(t)$ can be well approximated by a polynomial

through a Taylor expansion. We have

$$f_i(t) = \sum_{k=0}^{\infty} \frac{f_i^{(k)}(T/2)}{k!}(t - T/2)^k$$

$$= \sum_{k=0}^{K} \frac{f_i^{(k)}(T/2)}{k!}(t - T/2)^k + \mathcal{O}\left(\frac{\mathcal{C}}{2^K}\right), \quad \text{(C11)}$$

where in deriving the second line, we have used the fact that

$$\left| \frac{f_i^{(k)}(T/2)}{k!}(t - T/2)^k \right| \leq \mathcal{C}\left(\frac{\Gamma T}{2}\right)^k \leq \frac{\mathcal{C}}{2^k}. \quad \text{(C12)}$$

We then define

$$p_i^K(t) = \sum_{k=0}^{K} \frac{f_i^{(k)}(T/2)}{k!}(t - T/2)^k. \quad \text{(C13)}$$

This is a degree-K polynomial satisfying

$$|f_i(t) - p_i^K(t)| \leq \frac{\mathcal{C}}{2^K} \quad \text{(C14)}$$

for $t \in [0, T]$.

Following Ref. [48], which in turn relies on Ref. [59], we generate independent and identically distributed samples $t_1, t_2, \ldots, t_m$ from the Chebyshev distribution on $[0, T]$, specified by the probability density function $(1/\pi)(t(T - t))^{-1/2}$. We then generate N_s classical shadows for each $t_j, j = 1, 2, \ldots, m$. Therefore we are able to generate estimates y_{ij} such that

$$\mathbb{E}[y_{ij}] = f_i(t_j), \qquad \text{var}[y_{ij}] = \mathcal{O}\left(\frac{\|O_i\|_{\text{shadow}}^2}{N_s}\right). \quad \text{(C15)}$$

Therefore, with an appropriately chosen constant factor, we have

$$|y_{ij} - f_i(t_j)| \leq \mathcal{O}\left(\frac{\|O_i\|_{\text{shadow}}^2}{N_s}\right) \quad \text{(C16)}$$

with probability at least 2/3. This indicates that

$$|y_{ij} - p_i^K(t_j)| \leq \mathcal{O}\left(\frac{\|O_i\|_{\text{shadow}}^2}{N_s}\right) + \frac{\mathcal{C}}{2^K} \quad \text{(C17)}$$

with probability at least 2/3. By the Chernoff-Hoeffding theorem, the above inequality holds for a majority of the $j = 1, 2, \ldots, m$ with probability at least $1 - e^{-\Omega(m)}$.

Using the robust polynomial interpolation method proposed in Ref. [59], and as stated in Theorem E.1

of Ref. [48], we can construct polynomials $\hat{p}_i(t)$ from $\{(t_j, y_{ij})\}$ such that

$$|p_i^K(t) - \hat{p}_i(t)| \leq \mathcal{O}\left(\frac{\|O_i\|_{\text{shadow}}^2}{N_s} + \frac{\mathcal{C}}{2^K}\right) \quad \text{(C18)}$$

for all $t \in [0, T]$ with probability at least $1 - \delta'$, by choosing

$$m = \mathcal{O}(K \log(K\delta'^{-1})). \quad \text{(C19)}$$

This then leads to

$$|f_i(t) - \hat{p}_i(t)| \leq \mathcal{O}\left(\frac{\|O_i\|_{\text{shadow}}^2}{N_s} + \frac{\mathcal{C}}{2^K}\right) \quad \text{(C20)}$$

for all $t \in [0, T]$. Therefore, $\hat{p}_i(t)$ is a good uniform approximation of $f_i(t)$ for each i. Alternatively, we can also construct the polynomial approximation using the methods outlined in Appendix C 1 a to get a similar cost scaling.

The above method can fail in two scenarios: either the error bound (C17) fails to hold for a majority of times, or the sampled times fail to correctly capture the profile of the function. The former failure scenario has its probability bounded by $e^{-\Omega(m)}$ by the Chernoff-Hoeffding theorem, and the latter by δ' from the robust polynomial interpolation procedure. As a result, if we want to keep the total failure probability for a single $f_i(t)$ to be at most δ, then we only need

$$e^{-\Omega(m)} + \delta' \leq \delta. \quad \text{(C21)}$$

To this end, and taking into account Eq. (C19), it suffices to choose

$$m = \mathcal{O}(K \log(K\delta'^{-1})), \qquad \delta' = \delta/2. \quad \text{(C22)}$$

We want the uniform approximation error in Eq. (C20) to be upper bounded by ϵ. Therefore we can choose

$$N_s = \mathcal{O}\left(\frac{\max_i \|O_i\|_{\text{shadow}}^2}{\epsilon^2}\right), \qquad K = \mathcal{O}(\log(\mathcal{C}\epsilon^{-1})). \quad \text{(C23)}$$

Combining the above analysis, in particular Eqs. (C22) and (C23), the total number of classical shadows we need is

$$N_s \times m = \mathcal{O}\left(\frac{\max_i \|O_i\|_{\text{shadow}}^2}{\epsilon^2} \log(\mathcal{C}\epsilon^{-1})\right.$$

$$\left. \times \log\left(\frac{\log(\mathcal{C}\epsilon^{-1})}{\delta}\right)\right). \quad \text{(C24)}$$

We summarize the above analysis in the following lemma.

Lemma 4. Let $f_i(t) = \text{Tr}[\rho(t)O_i]$ for $i = 1, 2, \ldots, \chi$. We further assume that $|f_i^{(k)}(t)| \leq C\Gamma^k k!$. Then, for $T \leq 1/\Gamma$, we can construct polynomials $\hat{p}_i(t)$, $i = 1, 2, \ldots, \chi$, with degree up to $\mathcal{O}(\log(\mathcal{C}\epsilon^{-1}))$ such that

$$\Pr\left[\max_{t \in [0,T]} |\hat{p}_i(t) - f_i(t)| > \epsilon \right] < \delta, \qquad (C25)$$

using

$$\mathcal{O}\left(\frac{\max_i \|O_i\|_{\text{shadow}}^2}{\epsilon^2} \log(\mathcal{C}\epsilon^{-1}) \log\left(\frac{\log(\mathcal{C}\epsilon^{-1})}{\delta} \right) \right) \qquad (C26)$$

classical shadows of the time-evolved state $\rho(t)$.

c. Long-time interpolation

We now consider the case where T is not necessarily upper bounded by $1/\Gamma$. In this case we can simply partition the interval $[0, T]$ into segments, each of length at most $1/\Gamma$, and there are therefore ΓT such segments. We then use the algorithm described in Appendix C 1 b to generate a polynomial to approximate each $f_i(t)$ on each of the ΓT segments. Piecing these polynomials together we have a piecewise polynomial approximation $\hat{g}_i(t)$ that approximates $f_i(t)$ for all $t \in [0, T]$. The success probability of this procedure can be obtained via a union bound. We therefore arrive at the following theorem from Lemma 4.

Theorem 6. Let $f_i(t) = \text{Tr}[\rho(t)O_i]$ for $i = 1, 2, \ldots, \chi$. We further assume that $|f_i^{(k)}(t)| \leq C\Gamma^k k!$. Then, for $T > 0$, we can construct piecewise polynomial functions $\hat{g}_i(t)$, $i = 1, 2, \ldots, \chi$, with degrees up to $\mathcal{O}(\log(\mathcal{C}\epsilon^{-1}))$ on at most ΓT segments such that

$$\Pr\left[\max_{t \in [0,T]} |\hat{g}_i(t) - f_i(t)| > \epsilon \right] < \delta, \qquad (C27)$$

using

$$\mathcal{O}\left(\frac{\Gamma T \max_i \|O_i\|_{\text{shadow}}^2}{\epsilon^2} \log(\mathcal{C}\epsilon^{-1}) \log\left(\frac{\Gamma T \log(\mathcal{C}\epsilon^{-1})}{\delta} \right) \right) \qquad (C28)$$

classical shadows of the time-evolved state $\rho(t)$.

2. Testing conservation laws

For each $i = 1, 2, \ldots, \chi$, we define the time average $\bar{f}_i = (1/T) \int_0^T f_i(t) dt$. For each i, we want to test which of the two following hypotheses is true.

Hypothesis 1. For all $t \in [0, T], f_i(t) = \bar{f}_i$.

Hypothesis 2. There exists $t^* \in [0, T]$ such that $|f_i(t^*) - \bar{f}_i| \geq \epsilon$.

Unlike the usual statistical hypothesis testing situation, the two hypotheses we consider above are treated on an equal footing, and therefore we do not need to distinguish between the null hypothesis and the alternative hypothesis. This is possible because we are considering a promise decision problem.

With the piecewise polynomial approximations $\hat{g}_i(t)$ we have, we can easily distinguish the two cases. An $\epsilon/8$-uniform approximation ensures that

$$|\hat{g}_i(t) - f_i(t)| \leq \frac{\epsilon}{8}, \qquad \left| \frac{1}{T} \int_0^T \hat{g}_i(t) dt - \bar{f}_i \right| \leq \frac{\epsilon}{8}. \quad (C29)$$

Therefore, if Hypothesis 1 is true then we have

$$\left| \frac{1}{T} \int_0^T \hat{g}_i(t) dt - \hat{g}_i(t) \right| \leq \frac{\epsilon}{8} + \frac{\epsilon}{8} = \frac{\epsilon}{4}. \qquad (C30)$$

If Hypothesis 2 is true then we have

$$\left| \frac{1}{T} \int_0^T \hat{g}_i(t) dt - \hat{g}_i(t^*) \right|$$
$$\geq |\bar{f}_i - f_i(t^*)| - \left| \frac{1}{T} \int_0^T \hat{g}_i(t) dt - \bar{f}_i \right| - |f_i(t^*) - \hat{g}_i(t^*)|$$
$$\geq \frac{3\epsilon}{4}. \qquad (C31)$$

Therefore, the two hypotheses can be distinguished by the statistic $\max_{t \in [0,T]} |(1/T) \int_0^T \hat{g}_i(t) dt - \hat{g}_i(t)| \leq \epsilon/4$ or $\geq 3\epsilon/4$.

We can successfully distinguish between the two cases if we get an ϵ-uniform approximation for $f_i(t)$. Therefore we can use Theorem 6 to determine the cost of keeping the error probability below δ (which means that both the type-I and type-II error probabilities are below δ). We therefore have the following theorem.

Theorem 7. Let $f_i(t) = \text{Tr}[\rho(t)O_i]$ for $i = 1, 2, \ldots, \chi$. We further assume that $|f_i^{(k)}(t)| \leq C\Gamma^k k!$. Then, for $T > 0$, we can distinguish between Hypotheses 1 and 2 for each i correctly using

$$\mathcal{O}\left(\frac{\Gamma T \max_i \|O_i\|_{\text{shadow}}^2}{\epsilon^2} \log(\mathcal{C}\epsilon^{-1}) \log\left(\frac{\Gamma T \log(\mathcal{C}\epsilon^{-1})}{\delta} \right) \right) \qquad (C32)$$

classical shadows of the time-evolved state $\rho(t)$.

APPENDIX D: TESTING CONSERVATION LAWS FOR AN ENSEMBLE OF INITIAL STATES

In this appendix, we upper bound the generalization error given in Eq. (9) in the main text, which we restate

here:

$$\left| \mathbb{E}_{\rho \sim \mathcal{D}} d(O_i, \rho) - \frac{1}{N_I} \sum_{k=1}^{N_I} d(O_i, \rho_k) \right|. \qquad \text{(D1)}$$

Note that $|d(O, \rho)| \leq 2\|O\|$ by the definition of $d(O, \rho)$ given in Eq. (7). As a result, by Hoeffding's inequality, the above generalization error is at most ϵ' with probability at least

$$1 - 2 \exp\left(-\frac{N_I \epsilon'^2}{2\|O\|^2} \right). \qquad \text{(D2)}$$

This means that, in order to make the generalization error at most ϵ with probability at least $1 - \delta'$, we need

$$N_I = \mathcal{O}(\epsilon^{-2} \log(\delta'^{-1}) \|O\|^2). \qquad \text{(D3)}$$

APPENDIX E: QUERY COMPLEXITY LOWER BOUND FOR TESTING CONSERVATION LAWS FOR ALL INITIAL STATES

In this appendix, we consider the setting where the Hamiltonian e^{-iHt} on N qubits is provided through an oracle, and we want to show that testing whether H commutes with a simple observable can require $\Omega(2^{N/2})$ queries to the oracle in the worst case.

Our result is based on the lower bound for computing the OR function. In this setting, an 2^N-bit string $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{2^N - 1})$ is provided through an oracle U that satisfies

$$U|n\rangle = (-1)^{\mathbf{x}_n} |n\rangle \qquad \text{(E1)}$$

for $n = 0, 1, \ldots, 2^N - 1$. In order to compute $OR(\mathbf{x})$ [78], it is known that at least $\Omega(2^{N/2})$ queries to U are needed [61]. This query complexity lower bound still holds even if we constrain $\mathbf{x}$ to contain at most a single 1, which corresponds to the partial function setting discussed in the comment after Theorem 4.13 of Ref. [61].

Now we choose our Hamiltonian $H = U$ to be this oracle unitary U, which is incidentally also Hermitian. We further restrict to the case where $\mathbf{x}$ contains at most a single 1. We first implement e^{-iHt} using the oracle U itself. Because the eigenvalues of $H = U$ are ± 1, we can implement e^{-iHt} using only two queries of U for arbitrary t through phase kickback.

We now assume that an algorithm can do the following: given access to e^{-iHt} (acting on N qubits) for arbitrarily chosen t, it can distinguish the two cases

$$[H, X_1] = 0 \quad \text{or} \quad \|[H, X_1]\| \geq 1, \qquad \text{(E2)}$$

where X_1 is the Pauli-X operator on the first qubit. If the algorithm can accomplish the task with Q queries to e^{-iHt},

we next argue that it can compute $OR(\mathbf{x})$ using $2Q$ queries to U (because of the implementation of e^{-iHt} discussed in the previous paragraph), and in this way show that $Q = \Omega(2^{N/2})$.

Our argument goes as follows: if $OR(\mathbf{x}) = 1$ then there exists $0 \leq n^* \leq N - 1$ such that $\mathbf{x}_{n^*} = 1$, whereas $\mathbf{x}_n = 0$ for all other n because of our restriction of the domain of the OR function. Therefore,

$$H = U = I - 2|n^*\rangle\langle n^*|. \qquad \text{(E3)}$$

One can then compute

$$\|[H, X_1]\| = 2\|[|n^*\rangle\langle n^*|, X_1]\| = 2, \qquad \text{(E4)}$$

where we have used the fact that $X_1|n^*\rangle$ is orthogonal to $|n^*\rangle$. On the other hand, if $OR(\mathbf{x}) = 0$ then $H = U = I$, and, as a result, $[H, X_1] = 0$. Therefore, as long as we can distinguish the two cases in Eq. (E2), we are able to evaluate $OR(\mathbf{x})$. The above argument therefore leads us to the following theorem.

Theorem 8. Given access to e^{-iHt} (acting on N qubits) for arbitrarily chosen t as a black-box oracle, any algorithm that can distinguish between $[H, X_1] = 0$ and $\|[H, X_1]\| \geq 1$ with probability at least $2/3$, where X_1 is the Pauli-X operator on the first qubit, takes at least $\Omega(2^{N/2})$ queries to e^{-iHt} in the worst case.

APPENDIX F: TIME-DERIVATIVE BOUNDS FOR LOCAL OBSERVABLE EXPECTATION VALUES

We consider a general Hamiltonian of the form

$$H = \sum_{P \in \{I, X, Y, Z\}^N} \lambda_P P, \qquad \text{(F1)}$$

where $|\lambda_P| \leq 1$, and $P = \bigotimes_j P_j$ is a Pauli operator with components $P_j \in \{I, X, Y, Z\}$. We show that the local observable expectation values behave nicely as a function of time for a class of Hamiltonians. More precisely, consider a local observable O, which, by definition, is supported on $s = \mathcal{O}(1)$ adjacent qubits; we want to bound the high-order derivatives of

$$\langle O(t) \rangle = \text{Tr}[e^{iHt} O e^{-iHt} \rho]. \qquad \text{(F2)}$$

The Hamiltonians H we consider need to satisfy the following assumption.

Lemma 5. Let H in Eq. (F1) be a k-local Hamiltonian. We assume that, for each qubit j, $\sum_{P:P_j \neq I} |\lambda_P| \leq \Lambda$. Then

$$\left| \frac{d^\ell}{dt^\ell} \langle O(t) \rangle \right| = \mathcal{O}(\ell! (2\Lambda(k-1))^\ell \|O\|). \qquad \text{(F3)}$$

Proof. First, we observe that

$$\frac{d^\ell}{dt^\ell} \langle O(t) \rangle = \text{tr}[e^{iHt} \, \text{ad}_H^\ell(O) e^{-iHt} \rho], \qquad \text{(F4)}$$

where we recall the notation that $\text{ad}_A(B) := [A, B]$. Therefore, it suffices to prove that

$$\| \text{ad}_H^\ell(O) \| = \mathcal{O}(\ell!(2\Lambda(k-1))^\ell \|O\|). \qquad \text{(F5)}$$

We first expand $\text{ad}_H^\ell(O)$ using the expression for H in Eq. (F1):

Then, for each P^r, there must exist a qubit q_r such that both P^r and O_{r-1} act nontrivially on q_r, in order for $[P^\ell, \ldots [P^2, [P^1, O]] \ldots] \neq 0$. Consequently, we have

$$\sum_{P^1, P^2, \ldots, P^\ell} |\lambda_{P^1} \lambda_{P^2} \cdots \lambda_{P^\ell}| \|[P^\ell, \ldots [P^2, [P^1, O]] \ldots]\|$$

$$\leq \sum_{q_1 \in \text{supp}(O)} \sum_{P^1 : P^1_{q_1} \neq I} \sum_{q_2 \in \text{supp}(O_1)} \sum_{P^2 : P^2_{q_2} \neq I} \cdots \sum_{q_\ell \in \text{supp}(O_{\ell-1})} \sum_{P^\ell : P^\ell_{q_\ell} \neq I} |\lambda_{P^1} \lambda_{P^2} \cdots \lambda_{P^\ell}| \|O\|$$

$$= \|O\| \sum_{q_1 \in \text{supp}(O)} \sum_{P^1 : P^1_{q_1} \neq I} |\lambda_{P^1}| \sum_{q_2 \in \text{supp}(O_1)} \sum_{P^2 : P^2_{q_2} \neq I} |\lambda_{P^2}| \cdots \sum_{q_\ell \in \text{supp}(O_{\ell-1})} \sum_{P^\ell : P^\ell_{q_\ell} \neq I} |\lambda_{P^\ell}|. \qquad \text{(F8)}$$

Note that $\sum_{P^\ell : P^\ell_{q_\ell} \neq I} |\lambda_{P^\ell}| \leq \Lambda$ by assumption. As a result,

$$\sum_{q_\ell \in \text{supp}(O_{\ell-1})} \sum_{P^\ell : P^\ell_{q_\ell} \neq I} |\lambda_{P^\ell}| \leq |\text{supp}(O_{\ell-1})|\Lambda$$

$$\leq (s + (\ell - 1)(k-1))\Lambda, \quad \text{(F9)}$$

where we have used the fact that $|\text{supp}(O_{\ell-1})| \leq s + (\ell - 1)(k-1)$, which can be proved by induction on ℓ. Because of this, the right-hand side of Eq. (F8) can be upper bounded by

$$(s + (\ell - 1)(k-1))\Lambda \|O\|$$

$$\times \sum_{q_1 \in \text{supp}(O)} \sum_{P^1 : P^1_{q_1} \neq I} |\lambda_{P^1}| \sum_{q_2 \in \text{supp}(O_1)} \sum_{P^2 : P^2_{q_2} \neq I} |\lambda_{P^2}| \cdots$$

$$\times \sum_{q_{\ell-1} \in \text{supp}(O_{\ell-2})} \sum_{P^\ell : P^{\ell-1}_{q_{\ell-1}} \neq I} |\lambda_{P^{\ell-1}}|. \qquad \text{(F10)}$$

One can keep doing this ℓ times, and the right-hand side of Eq. (F8) is bounded by

$$\|O\|\Lambda^\ell s(s + k - 1) \cdots (s + (\ell - 1)(k-1)). \qquad \text{(F11)}$$

Because of Eq. (F6), this is an upper bound of $\| \text{ad}_H^\ell(O) \|$.

$$\text{ad}_H^\ell(O) = \sum_{P^1, P^2, \ldots, P^\ell} \lambda_{P^1} \lambda_{P^2} \cdots \lambda_{P^\ell}$$

$$\times [P^\ell, \ldots [P^2, [P^1, O]] \ldots]. \qquad \text{(F6)}$$

Note that the nested commutator $[P^\ell, \ldots [P^2, [P^1, O]] \ldots] \neq 0$ only if each P^r overlaps with the nested commutator up to the $r - 1$ level, $r = 1, 2, \ldots, \ell$. For a fixed sequence of $\{P^r\}$, we recursively define these nested commutators through

$$O_0 = O, \qquad O_r = [P^r, O_{r-1}], \quad r = 1, 2, \ldots, \ell. \quad \text{(F7)}$$

We only need to bound $s(s + k - 1) \cdots (s + (\ell - 1)(k-1))$. We have

$$s(s + k - 1) \cdots (s + (\ell - 1)(k-1))$$

$$= (k-1)^\ell \frac{s}{k-1}\left(\frac{s}{k-1} + 1\right) \cdots \left(\frac{s}{k-1} + \ell - 1\right)$$

$$\leq (k-1)^\ell \frac{(\lceil s/(k-1) \rceil + \ell - 1)!}{\lceil s/(k-1) - 1 \rceil!}$$

$$= (k-1)^\ell \binom{\lceil s/(k-1) \rceil + \ell - 1}{\ell} \ell!$$

$$\leq (k-1)^\ell 2^{\lceil s/(k-1) \rceil + \ell - 1} \ell!$$

$$= \mathcal{O}((2(k-1))^\ell \ell!). \qquad \text{(F12)}$$

This completes the proof. ∎

Next, we show that geometrically local Hamiltonians and local Hamiltonians with power-law interaction that decays fast enough satisfy the assumptions of Lemma 5. The key quantity of interest is $\sum_{P : P_j \neq I} |\lambda_P|$, which is the sum of the absolute value of all coefficients of Pauli terms that act nontrivially on a qubit j. For geometrically local Hamiltonians, there are only $\mathcal{O}(1)$ terms acting on any given qubit, and, consequently, $\sum_{P : P_j \neq I} |\lambda_P| = \mathcal{O}(1)$ if $|\lambda_P| \leq 1$, as assumed at the beginning of this section.

For power-law interaction Hamiltonians, we adopt a restricted definition to make the discussion easier, without neglecting any essential feature of these Hamiltonians. For these Hamiltonians, $\lambda_P \neq 0$ only when P involves at most two qubits. Moreover, $|\lambda_P| = \mathcal{O}(d^{-\alpha})$, where d is the distance, on a D-dimensional lattice, between the two qubits, and α is the exponent deciding how rapid the decay is. The sum of all coefficients involving a qubit j can be roughly bounded by

$$\sum_{j' \in \mathbb{Z}^D} |j'|^{-\alpha}, \tag{F13}$$

where Z is the set of all integers, and $\mathbb{Z}^D$ is a D-dimensional lattice. When $\alpha > D + 1$, we have $\sum_{j' \in \mathbb{Z}^D} |j'|^{-\alpha} < \infty$, thus giving us a bound $\sum_{P:P_j \neq I} |\lambda_P| = \mathcal{O}(1)$.

----

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[2] S.-M. Udrescu and M. Tegmark, AI Feynman: A physics-inspired method for symbolic regression, Sci. Adv. **6**, eaay2631 (2020).

[3] B. M. de Silva, D. M. Higdon, S. L. Brunton, and J. N. Kutz, Discovery of physics from data: Universal laws and discrepancies, Front. Artif. Intell. **3** (2020).

[4] Z. Liu, V. Madhavan, and M. Tegmark, Machine learning conservation laws from differential equations, Phys. Rev. E **106**, 045307 (2022).

[5] P. Y. Lu, R. Dangovski, and M. Soljačić, Discovering conservation laws using optimal transport and manifold learning, Nat. Commun. **14**, 4744 (2023).

[6] R. D. King, J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, and A. Clare, The automation of science, Science **324**, 85 (2009).

[7] M. Schmidt and H. Lipson, Distilling free-form natural laws from experimental data, Science **324**, 81 (2009).

[8] T. Wu and M. Tegmark, Toward an artificial intelligence physicist for unsupervised learning, Phys. Rev. E **100**, 033311 (2019).

[9] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, Proc. Natl. Acad. Sci. **116**, 22445 (2019).

[10] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, Discovering physical concepts with neural networks, Phys. Rev. Lett. **124**, 010508 (2020).

[11] Y.-I. Mototake, Interpretable conservation law estimation by deriving the symmetries of dynamics from trained deep neural networks, Phys. Rev. E **103**, 033303 (2021).

[12] S. J. Wetzel, R. G. Melko, J. Scott, M. Panju, and V. Ganesh, Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks, Phys. Rev. Res. **2**, 033499 (2020).

[13] F. Alet, D. Doblar, A. Zhou, J. Tenenbaum, K. Kawaguchi, and C. Finn, in *Advances in Neural Information Processing Systems*, Vol. 34, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (Curran Associates, Inc., 2021), p. 16384.

[14] E. Kaiser, J. N. Kutz, and S. L. Brunton, in *2018 IEEE Conference on Decision and Control (CDC)* (IEEE, Miami Beach, FL, 2018), p. 6415.

[15] Z. Liu and M. Tegmark, Machine learning conservation laws from trajectories, Phys. Rev. Lett. **126**, 180604 (2021).

[16] S. Ha and H. Jeong, Discovering invariants via machine learning, Phys. Rev. Res. **3**, L042035 (2021).

[17] P. Calabrese, F. H. L. Essler, and G. Mussardo, Introduction to "quantum integrability in out of equilibrium systems", J. Stat. Mech.: Theory Exp. **2016**, 064001 (2016).

[18] M. Serbyn, Z. Papić, and D. A. Abanin, Local conservation laws and the structure of the many-body localized states, Phys. Rev. Lett. **111**, 127201 (2013).

[19] D. A. Huse, R. Nandkishore, and V. Oganesyan, Phenomenology of fully many-body-localized systems, Phys. Rev. B **90**, 174202 (2014).

[20] V. Ros, M. Müller, and A. Scardicchio, Integrals of motion in the many-body localized phase, Nucl. Phys. B **891**, 420 (2015).

[21] J. Z. Imbrie, V. Ros, and A. Scardicchio, Local integrals of motion in many-body localized systems, Ann. Phys. **529**, 1600278 (2017).

[22] P. W. Anderson, Absence of diffusion in certain random lattices, Phys. Rev. **109**, 1492 (1958).

[23] L. Fleishman and P. W. Anderson, Interactions and the Anderson transition, Phys. Rev. B **21**, 2366 (1980).

[24] I. V. Gornyi, A. D. Mirlin, and D. G. Polyakov, Interacting electrons in disordered wires: Anderson localization and low-t transport, Phys. Rev. Lett. **95**, 206603 (2005).

[25] D. Basko, I. Aleiner, and B. Altshuler, Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states, Ann. Phys. (N. Y.) **321**, 1126 (2006).

[26] D. A. Abanin and Z. Papić, Recent progress in many-body localization, Ann. Phys. **529**, 1700169 (2017).

[27] R. Nandkishore and D. A. Huse, Many-body localization and thermalization in quantum statistical mechanics, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).

[28] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Colloquium: Many-body localization, thermalization, and entanglement, Rev. Mod. Phys. **91**, 021001 (2019).

[29] S. Moudgalya, B. A. Bernevig, and N. Regnault, Quantum many-body scars and Hilbert space fragmentation: A review of exact results, Rep. Progr. Phys. **85**, 086501 (2022).

[30] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[31] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2022).

[32] A. Chandran, I. H. Kim, G. Vidal, and D. A. Abanin, Constructing local integrals of motion in the many-body localized phase, Phys. Rev. B **91**, 085425 (2015).

[33] M. Mierzejewski, P. Prelovšek, and T. C. V. Prosen, Identifying local and quasilocal conserved quantities in integrable systems, Phys. Rev. Lett. **114**, 140601 (2015).

[34] T. E. O'Brien, D. A. Abanin, G. Vidal, and Z. Papić, Explicit construction of local conserved operators in disordered many-body systems, Phys. Rev. B **94**, 144208 (2016).

[35] E. Chertkov, B. Villalonga, and B. K. Clark, Numerical evidence for many-body localization in two and three dimensions, Phys. Rev. Lett. **126**, 180602 (2021).

[36] G. Bentsen, I.-D. Potirniche, V. B. Bulchandani, T. Scaffidi, X. Cao, X.-L. Qi, M. Schleier-Smith, and E. Altman, Integrable and chaotic dynamics of spins coupled to an optical cavity, Phys. Rev. X **9**, 041011 (2019).

[37] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, and A. Laing, *et al.*, Experimental quantum Hamiltonian learning, Nat. Phys. **13**, 551 (2017).

[38] T. J. Evans, R. Harper, and S. T. Flammia, Scalable Bayesian Hamiltonian learning, ArXiv:1912.07636.

[39] A. Gu, L. Cincio, and P. J. Coles, Practical black box Hamiltonian learning, ArXiv:2206.15464.

[40] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online Hamiltonian learning, New J. Phys. **14**, 103013 (2012).

[41] D. Hangleiter, I. Roth, J. Eisert, and P. Roushan, Precise Hamiltonian identification of a superconducting quantum processor, ArXiv:2108.08319 (2021).

[42] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, Quantum Hamiltonian learning using imperfect quantum resources, Phys. Rev. A **89**, 042314 (2014).

[43] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Hamiltonian learning and certification using quantum resources, Phys. Rev. Lett. **112**, 190501 (2014).

[44] W. Yu, J. Sun, Z. Han, and X. Yuan, Practical and efficient Hamiltonian learning, ArXiv:2201.00190 (2022).

[45] A. Zubida, E. Yitzhaki, N. H. Lindner, and E. Bairey, Optimal short-time measurements for Hamiltonian learning, ArXiv:2108.08824.

[46] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, Robust online Hamiltonian learning, New J. Phys. **14**, 103013 (2012).

[47] Z. Li, L. Zou, and T. H. Hsieh, Hamiltonian tomography via quantum quench, Phys. Rev. Lett. **124**, 160502 (2020).

[48] D. S. Franca, L. A. Markovich, V. Dobrovitski, A. H. Werner, and J. Borregaard, Efficient and robust estimation of many-qubit Hamiltonians, ArXiv:2205.09567.

[49] J. Haah, R. Kothari, and E. Tang, Optimal learning of quantum Hamiltonians from high-temperature Gibbs states, ArXiv:2108.04842 (2021).

[50] L. Pastori, T. Olsacher, C. Kokail, and P. Zoller, Characterization and verification of Trotterized digital quantum simulation via Hamiltonian and Liouvillian learning, PRX Quantum **3**, 030324 (2022).

[51] M. C. Caro, Learning quantum processes and Hamiltonians via the Pauli transfer matrix, ArXiv:2212.04471.

[52] H.-Y. Huang, Y. Tong, D. Fang, and Y. Su, Learning many-body Hamiltonians with Heisenberg-limited scaling, ArXiv:2210.03030.

[53] T.-L. Zhao, S.-X. Hu, and Y. Zhang, Maximum-likelihood-estimate Hamiltonian learning via efficient and robust quantum likelihood gradient, Phys. Rev. Res. **5**, 023136 (2023).

[54] O. Shtanko, D. S. Wang, H. Zhang, N. Harle, A. Seif, R. Movassagh, and Z. Minev, Uncovering local integrability in quantum many-body dynamics, ArXiv:2307.07552.

[55] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).

[56] D. E. Koh and S. Grewal, Classical shadows with noise, Quantum **6**, 776 (2022).

[57] V. Vitale, A. Rath, P. Jurcevic, A. Elben, C. Branciard, and B. Vermersch, Estimation of the quantum Fisher information on a quantum processor, ArXiv:2307.16882 (2023).

[58] M. L. LaBorde and M. M. Wilde, Quantum algorithms for testing Hamiltonian symmetry, Phys. Rev. Lett. **129**, 160503 (2022).

[59] D. Kane, S. Karmalkar, and E. Price, in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Berkeley, CA, 2017), p. 391.

[60] We use the asymptotic notation $\tilde{\mathcal{O}}(f(x))$ to denote $\mathcal{O}(f(x)\text{polylog}(f(x)))$.

[61] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. De Wolf, Quantum lower bounds by polynomials, J. ACM (JACM) **48**, 778 (2001).

[62] A. Pal and D. A. Huse, Many-body localization phase transition, Phys. Rev. B **82**, 174411 (2010).

[63] M. Žnidarič, T. Prosen, and P. Prelovšek, Many-body localization in the Heisenberg *XXZ* magnet in a random field, Phys. Rev. B **77**, 064426 (2008).

[64] D. J. Luitz, N. Laflorencie, and F. Alet, Many-body localization edge in the random-field Heisenberg chain, Phys. Rev. B **91**, 081103 (2015).

[65] The precise transition disorder strength in the thermodynamic limit is subject to ongoing research; see, e.g., Ref. [79] and the references therein.

[66] J. M. Deutsch, Quantum statistical mechanics in a closed system, Phys. Rev. A **43**, 2046 (1991).

[67] M. Srednicki, Chaos and quantum thermalization, Phys. Rev. E **50**, 888 (1994).

[68] In practice, we reconstruct all Pauli expectation values at a given time from the same experimental randomized measurement data. Thus, shot noise on different expectation values is in principle correlated. For a large number of measurements, however, we expect these correlations to be weak, and indeed confirm numerically that results obtained using the Gaussian noise approximation and actual randomized measurements are consistent.

[69] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, Random quantum circuits, Annu. Rev. Condens. Matter Phys. **14**, 335 (2023).

[70] V. V. Albert, Asymptotics of quantum channels: Conserved quantities, an adiabatic limit, and matrix product states, Quantum **3**, 151 (2019).

[71] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, ArXiv:2210.00921 (2023).

[72] A. Elben, B. Vermersch, M. Dalmonte, J. Cirac, and P. Zoller, Rényi entropies from random quenches in atomic Hubbard and spin models, Phys. Rev. Lett. **120**, 050406 (2018).

[73] J. Bringewatt, J. Kunjummen, and N. Mueller, Randomized measurement protocols for lattice gauge theories, ArXiv:2303.15519 (2023).

[74] L. Mirsky, Symmetric gauge functions and unitarily invariant norms, Q. J. Math. **11,** 50 (1960).

[75] G. Pisier, *Introduction to Operator Space Theory* (Cambridge University Press, 2003), Vol. 294.

[76] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel, Matrix concentration inequalities and free probability, ArXiv:2108.06312 (2021).

[77] L. N. Trefethen, *Spectral Methods in MATLAB* (SIAM, Philadelphia, 2000).

[78] Here, $OR(\mathbf{x})$ is defined as $OR(\mathbf{x}) = 0$ if all the $\mathbf{x}_i$ are 0 and $OR(\mathbf{x}) = 1$ otherwise.

[79] D. M. Long, P. J. D. Crowley, V. Khemani, and A. Chandran, Phenomenology of the prethermal many-body localized regime, ArXiv:2207.05761.