TU Delft

Delft University of Technology

Document Version
Final published version

Citation (APA)
Valenti, A., Jin, G., Léonard, J., Huber, S. D., & Greplova, E. (2022). Scalable Hamiltonian learning for large-scale out-of-equilibrium quantum dynamics. *Physical Review A*, *105*(2), Article 023302. https://doi.org/10.1103/PhysRevA.105.023302

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Scalable Hamiltonian learning for large-scale out-of-equilibrium quantum dynamics

Agnes Valenti,[1,*] Guliuxin Jin,[2,1,*] Julian Léonard,[3] Sebastian D. Huber,[1] and Eliska Greplova[2,1]
[1]*Institute for Theoretical Physics, ETH Zurich, CH-8093, Switzerland*
[2]*Kavli Institute of Nanoscience, Delft University of Technology, Delft 2628 CJ, Netherlands*
[3]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 18 March 2021; accepted 23 December 2021; published 1 February 2022)

Large-scale quantum devices provide insights beyond the reach of classical simulations. However, for a reliable and verifiable quantum simulation, the building blocks of the quantum device require exquisite benchmarking. This benchmarking of large-scale dynamical quantum systems represents a major challenge due to lack of efficient tools for their simulation. Here, we present a scalable algorithm based on neural networks for Hamiltonian tomography in out-of-equilibrium quantum systems. We illustrate our approach using a model for a forefront quantum simulation platform: ultracold atoms in optical lattices. Specifically, we show that our algorithm is able to reconstruct the Hamiltonian of an arbitrary sized bosonic ladder system using an accessible amount of experimental measurements. We are able to significantly increase the previously known parameter precision.

DOI: 10.1103/PhysRevA.105.023302

I. INTRODUCTION

Quantum simulators are at the forefront of quantum technologies and allow for the simulation of complex many-body Hamiltonians using direct control and manipulation of existing quantum systems. A quantum simulation is generally tailored towards studying a specific type of phenomena, typically one which is hard to simulate numerically on a classical computer. One particularly relevant example of such phenomena is out-of-equilibrium physics. Quantum simulations have been successfully implemented in a range of platforms: trapped ions [1–4], superconducting qubits [5–8], semiconductor quantum dots [9], and ultracold atoms [10–17]. Many of these currently available experimental systems are reaching sizes that are prohibitive for their classical exact simulation [3,11]. This fact raises a challenge of the *verification of quantum simulators*: While it is true that quantum simulators are able to address the issue of exploring physics that is intractable otherwise, the experimental control has intrinsic precision limits that lead to both errors and finite precision of the quantum simulation. We therefore need to develop tools to verify the performance of quantum simulations to ensure that the correct physics is implemented.

Recent progress has been made in reconstructing generic local Hamiltonians from measurements on single eigenstates and steady states of closed and open systems [18–27] as well as in the development of more specialized approaches tailored to concrete models [28–35]. An area of research that poses specific challenges for numerical methods and thus Hamiltonian tomography is out-of-equilibrium physics [36–39]. While many ground states and steady states of quantum systems can be readily captured by a range of approximate methods [40–42], out-of-equilibrium behavior of quantum systems proved to be significantly more challenging. Indeed, the verification of large-scale quantum out-of-equilibrium dynamics provide a readily available path towards quantum computational advantage [43,44].

An additional difficulty in the verification of the implemented underlying Hamiltonian arises in the presence of imperfect measurement outcomes caused by statistical averaging or measurement errors. Machine-learning methods have been shown to provide reliable predictions in the presence of experimental noise, even when trained on theoretical data [30,45]. Furthermore, the short timescale and low computational cost of evaluating a pretrained neural network model compared with more traditional methods such as Bayesian parameter estimation is compatible with most relevant experimental timescales.

In the present work, we introduce a scalable method to learn the Hamiltonian governing out-of-equilibrium quantum simulations. We develop a neural-network-based Hamiltonian learning technique that enables the reconstruction of the dynamics of large-scale quantum systems from experimentally accessible measurements with ultrahigh precision at low computational cost. We exemplify our approach by learning Hamiltonian parameters governing the dynamics of quenched ultracold bosons in a ladder geometry (quasi one dimensional) of optical lattice.

This paper is organized as follows: In Sec. II we present our algorithm and exemplify its practical implementation on an out-of-equilibrium bosonic system on a ladder. In Sec. III we present results of the Hamiltonian reconstruction for experimentally relevant parameter regimes. In Sec. IV we compare with a minimal Bayesian estimation benchmark in order to contextualize the obtained estimation errors. In Sec. V we address how to experimentally scale our approach to arbitrary lattice sizes. Finally, in Sec. VI we present the discussion and the outlook of our results.

[*]These authors contributed equally to this work.

 ©2022 American Physical Society

II. HAMILTONIAN LEARNING

The central goal of our method is to reconstruct the Hamiltonian governing the dynamics of a quantum system, or, in other words, to find a set of parameters required to fully characterize the dynamics of the system under consideration. Additionally, we wish to reconstruct the Hamiltonian with a maximum possible accuracy from a practically accessible amount of experimental measurements. Our protocol relies on postprocessing the measurement outcomes and using the so-obtained accessible and efficient representation of the relevant information to train and evaluate neural networks for the parameter estimation.

Many practically and experimentally relevant Hamiltonians have a local structure, i.e., they are sums of terms that only act on sites with a finite separation. This structure does not prevent long-range correlations to arise in the system, but it largely simplifies their verification: Since our goal is to reconstruct Hamiltonians, not wave functions, we may take advantage of this local structure and reconstruct the Hamiltonians from studying the physics of a subsystem of the experimentally relevant system. We show below that Hilbert spaces of these local subsystems become nearly or completely numerically manageable. Additionally, while the associated parameter space for the Hamiltonian of interest is not tractable, we show how to effectively approximate it in a scalable manner.

We explain our method on a concrete example of out-of-equilibrium quantum simulations with bosons in an optical lattice on a ladder. The neutral bosonic atoms are trapped in an optical lattice formed by laser beams through the coupling of the dipole moment of the atom with an incoming light [46,47]. This system has recently emerged as an excellent platform to explore out-of-equilibrium physics [13,14].

We study the Bose-Hubbard Hamiltonian [46,48] on a ladder,

$$H_{\mathrm{BH}} = -\sum_{\langle ij\rangle} J_{ij}\hat{a}_i^\dagger \hat{a}_j + \sum_i \frac{U_i}{2}\hat{a}_i^\dagger \hat{a}_i\left(\hat{a}_i^\dagger \hat{a}_i - 1\right) - \sum_i \mu_i \hat{a}_i^\dagger \hat{a}_i, \tag{1}$$

where the first term corresponds to particle hopping between neighboring sites and the coefficients J_{ij} correspond to the hopping amplitudes between site i and site j. The second term corresponds to on-site interaction and the coefficients U_i denote the strength of this interaction (in our case repulsion). The third term represents an on-site energy μ_i. We model here the main source of imperfection as offsets in the hopping parameters, on-site repulsions, and chemical potentials. As a consequence, we are now tasked with determining the most probable values of the parameters $\theta = \{J_{ij}, U_i, \mu_i\}$ conditioned on the outcome of obtained measurements. Following state-of-the art experiments with ultracold atoms, other decoherence sources and imperfections are strongly suppressed with respect to the modeled offsets in the Bose-Hubbard Hamiltonian [14]. We provide a more detailed explanation of the sources of imperfections in Appendix A.

Since the Hamiltonian in Eq. (1) has a local structure where the operators act at most on a pair of neighboring sites, we consider whether we can reconstruct all the required parameters by measurements on a smaller, more tractable subsystem.

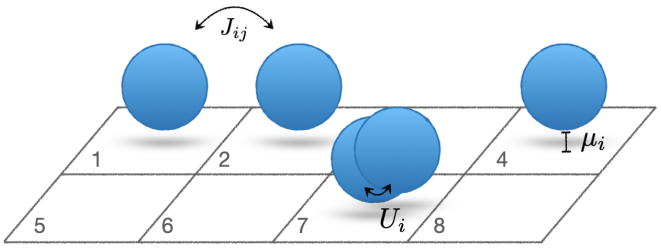


FIG. 1. Schematic plot of a subsystem of a ladder (quasi one dimensional) optical lattice with bosons. Each square corresponds to a lattice site with the system parameters indicated. This system constitutes the building block of our estimation algorithm.

Additionally, we expect the parameters to be reconstructed with higher precision on a smaller subsystem because the number of parameters dictating the system's behavior and thus the measurement outcomes is reduced. We emphasize here that we aim to estimate parameters of the Hamiltonian, not the wave function of the quantum system. As a consequence, while quantum correlation patterns in the wave function of a large quantum system can be qualitatively different than those of small subsystems, the Hamiltonian of the large system is still built by local terms. Thus, the validity of the estimation scheme presented here relies on the locality of the Hamiltonian and not on the wave-function correlations. In an experimental setup with ultracold atoms, it is very well suitable to create smaller subsystems by locally blocking hopping between specific sites and thus isolating a number of selected particles into the preselected area of the lattice [49]. In particular, projecting high-resolution optical potentials with holographic beam shaping has become an established technique in quantum gas experiments [50]. Since the optical potentials routinely reach a precision of $<10^{-4}$, it is possible to create box potentials with negligible influence on the system site potential offsets. More concretely, the parameters J_{ij}, U_i, and μ_i, sufficiently distant from the border, will reflect their values in the extended system. As a consequence, one may faithfully estimate the parameters of the global Hamiltonian by measurements solely on the experimentally created subsystems for this setting.

We consider an experimental setup similar to [13,14], consisting of a $2\times L$ ladder geometry, with L being the system length. We then propose to isolate and examine a 2×4 sites subsystem, as shown in Fig. 1. For four bosons, the Hilbert space size is 330 and thus the system is tractable via exact diagonalization. We use this tractable system of eight wells with four particles as the basic building block for the training of our algorithm. We consider the following out-of-equilibrium experimental sequence: the system is initialized in an easily prepared state with a trivial Hamiltonian followed by a quench of the Hamiltonian, i.e., a rapid change of the system parameters out of the Mott-insulating phase. After a time evolution for a time T we perform a set of measurements on the system. While the specific measurement performed depends on the concrete experiment, in the case of cold atoms we have access to quantum gas microscopy [51,52] that allows for direct visualization of the occupation number in each well of the optical lattice. In other words, we have the ability to project

onto the number states of the system after the conducted time evolution. One such Fock-state projection we denote as a "measurement snapshot." A set of measurement snapshots, corresponding to effective Monte Carlo sampling of the state, is obtained by repeatedly preparing the same initial state and performing the time evolution for the time T. A concrete possible experimental sequence is detailed in Refs. [13,14]. The practically achievable number of snapshots is on the order of 10^4, introducing a statistical error to all inferred quantities and the reconstruction of the output probability distribution [13,14]. From these measurements we want to infer the Hamiltonian governing the time evolution as precisely as possible. In particular, we assume that the parameters of the Bose-Hubbard Hamiltonian are unknown and we aim to estimate the values of the parameters as precisely as possible. As a consequence of the restriction on the amount of experimentally feasible measurements, we not only want to estimate the Hamiltonian with the maximum possible precision, but at the same time we need to be able to achieve that with as little data as possible.

We note here that a uniform chemical potential only yields a global phase factor and is thus not detectable within the closed system with a constant number of atoms. Instead, we reconstruct the fluctuations of the chemical potentials by considering the quantities $\mu_{\mathrm{diff},i} := \mu_i - \mu_1, i \geqslant 2$. Thus, we aim to estimate the set of in total 25 parameters $\{J_{ij}, U_i, \mu_{\mathrm{diff},i}\}$ fully characterizing the system's dynamics.

In the following we explain how to design a computationally efficient machine-learning algorithm for the Hamiltonian reconstruction. Additionally, we use the well-established but computationally costly Bayesian parameter estimation [53,54] as a benchmark to assess the quality of the presented neural-network-based Hamiltonian tomography process. Neural networks are, in general, capable of approximating an arbitrary function and once the network is trained it is possible to evaluate it on new input at low computational cost. Recent results suggest that classical machine learning is in general powerful in the analysis of quantum systems [55–59]. Our goal is to approximate the map from postprocessed experimental measurements M to the set of parameters that fully specify the Hamiltonian.

Neural networks are trained using datasets of example data instances, the more examples the network sees the more general models we are able to build. Additionally, accurate results are more readily obtained when the function we wish to approximate has a low level of complexity. To reduce the complexity of the function mapping the measurement outcomes to the respective Hamiltonian parameters, we take two measures: First we postprocess the measurement outcomes such that the relevant physics are encoded in a more accessible manner. As the postprocessing represents a compression of the measurement data, it yields the additional advantage of significantly reducing the training set size resulting in a tractable computational training cost. Second, we factorize the problem by training a separate model for each parameter we wish to estimate.

The postprocessing of the measurement data is based on calculating a set of relevant density correlators capturing the necessary information about the set of Bose-Hubbard parameters. Specifically we calculate the average occupation number $\langle n_i \rangle$, $i = 1, \ldots, 8$ for each well. Given that the densities are particularly sensitive to the on-site repulsions and fluctuations of the chemical potentials, relevant information for the estimation of those parameters is compressed here. Correlators of the density between any two $\langle n_i n_j \rangle$, three $\langle n_i n_j n_k \rangle$, and four $\langle n_i n_j n_k n_l \rangle$ wells yield more complete information because they are also highly influenced by the values of the hopping amplitudes between wells. Additionally, we calculate the correlators $\langle n_i(n_i - 1) \rangle$ in order to emphasize multiple occupancies. Given that we are using four particles in eight wells, any higher-order correlations are negligibly small. This construction results in 171 correlation values that can be organized into a one-dimensional (1D) vector $\vec{n}$, which presents a much more compact representation of the original measurement dataset. We provide detailed information on the construction of the vector $\vec{n}$ in Appendix C.

The factorization of the classification problem is performed as follows: instead of building a single model that predicts a vector of 25 parameters for each measured dataset, we choose to build separate models for each parameter, i.e., 25 models in total.

We use 25 feed-forward neural networks trained using supervised learning. In all instances the input is the correlator vector $\vec{n}$ and the output is a single real number predicting the value of the given parameter. The networks consist of one input layer with 171 neurons (one neuron for each expectation value) and five fully connected hidden layers with 300, 400, 300, 150, and 100 neurons, respectively. To train the networks, we create 150 500 training examples. Each of these training examples consists of a pair of inputs and labels J^{label}, U^{label}, $\mu^{\mathrm{label}}_{\mathrm{diff}}$, where the labels ("true" values) correspond to one parameter configuration chosen randomly within the experimental uncertainty. The input vector $\vec{n}$ is obtained by using exact diagonalization to simulate the time evolution and taking measurement snapshots of the system. The time evolution and initial state of the simulated system can be chosen as experimentally feasible. In particular, we choose an initial state with four atoms with an atom on every second site and the time evolution $T = 200/\langle J_{ij} \rangle$. Testing with smaller T did not show any significant difference in precision, therefore the neural network may be trained for a specific experiment by using the most experimentally suitable time evolution T. The correlators $\vec{n}$ are then calculated as expectation values of the simulated measurements. The training examples are used to train 25 networks minimizing the objective cost function

$$C(X) = |X^{\mathrm{pred}} - X^{\mathrm{label}}|^2. \tag{2}$$

Here, we use X as placeholder for the parameter to be estimated. This choice of cost function allows for accurate estimation of continuous parameters, representing a large advantage to discrete, more conventional such methods as standard Bayesian parameter estimation models. We provide further detailed information on the architecture and training in Appendix C.

III. RESULTS

The trained networks can then be applied to estimate parameters in an experiment. We simulate the experimental sequence by using exact diagonalization. We show the

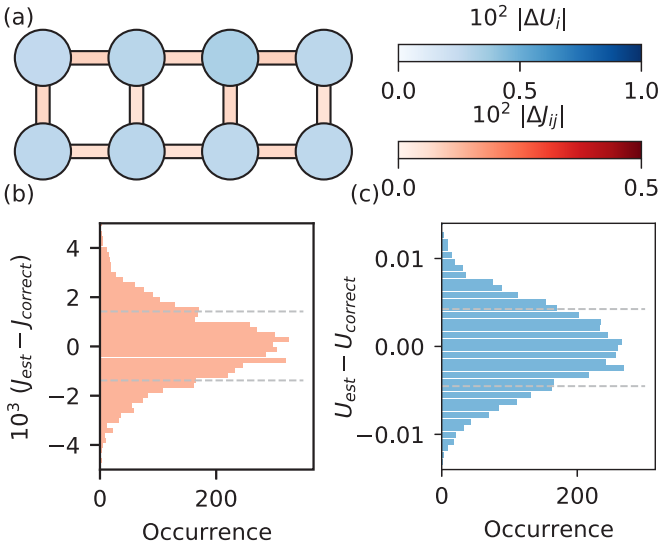


FIG. 2. Precision of the neural-network estimation scheme for 2500 measurement snapshots: Panel (a) shows the size of the estimation errors regarding the respective positions of the on-site repulsion (blue circles) and the neighboring site hopping (orange connections). Panel (b) [(c)] shows the error distribution independent of the spatial location for the hopping amplitudes (on-site repulsion) parameters over 500 parameter configurations $\{J_{ij}, U_i, \mu_i\}$, each evaluated using 2500 snapshots. Dashed gray lines denote the standard deviation of the error.

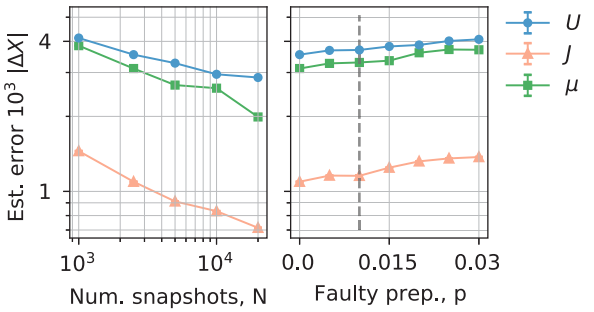


FIG. 3. Absolute estimation error as a function of number of measurement snapshots (left) and as a function of the probability p to not place an atom on the intended site during the preparation of the initial state (right). The estimation error is taken for neighboring sites hopping amplitudes (orange triangles), on-site repulsion (blue circles), and chemical potential differences (green squares). The average is taken over a test set of 500 data sets, each set corresponding to a parameter configuration $\{J_{ij}, U_i, \mu_i\}$.

results of the neural-network estimation in Fig. 2 for the case of 2500 simulated measurement snapshots. Figure 2(a) shows the average errors for the hopping amplitudes $\Delta J_{ij} = |J_{ij}^{\text{correct}} - J_{ij}^{\text{estimated}}|$ and the on-site repulsion potentials $\Delta U_i = |U_i^{\text{correct}} - U_i^{\text{estimated}}|$ in their spatial location. We denote here as "correct" parameters the parameters used to simulate the system, which are typically unknown in an experiment. In particular, here we simulate those experimental conditions by choosing a set of correct parameters J_{ij}^{correct}, U_i^{correct}, and μ_i^{correct} within an interval of 1% (±0.5%) around the average values $J \approx 1$, $U \approx 2$, and $\mu \approx 1$. The absolute uncertainty interval of the chemical-potential differences $\mu_{\text{diff},i}$ is by construction twice the uncertainty of the chemical potentials μ_i. The blue circles in Fig. 2 represent on-site repulsion errors and their red connections represent the intersite hopping errors of the neural-network estimation. Figures 2(b) and 2(c) show the distribution of the errors for hopping amplitudes and on-site repulsion, respectively. Gray dashed lines indicate the standard deviation of the distributions. The shown data have been averaged over the test set consisting of 500 measurement sets (2500 measurement snapshots each). We observe that the absolute error is around 0.1×10^{-2} for hopping amplitudes, 0.35×10^{-2} for the on-site repulsion and 0.3×10^{-2} for the chemical-potential differences, therefore significantly improving over the prior precision for all parameters. We elaborate in Appendix E on the dependence of the estimated precision with respect to the experimentally known uncertainty.

The precision of the estimation depends on the accuracy of the computed correlator expectation values. The main source of error is the statistical error arising from a finite amount of measurement snapshots, as discussed above. In particular, we have shown the precision for 2500 measurement snapshots in Fig. 2, i.e., a relatively low number. We therefore examine the scaling of the absolute error ΔX with $X \in \{J, U, \mu_{\text{diff}}\}$ as a function of measured snapshots with the error averaged over all hopping amplitudes, on-site repulsions, and chemical-potential differences in Fig. 3. More concretely, examining the errors ΔX for the hopping potentials, on-site repulsions, and chemical-potential differences, we see that, in the most experimentally demanding case of 20 000 snapshots, the averaged estimated absolute error is $<0.1\times10^{-2}$, $<0.3\times10^{-2}$, and $<0.25\times10^{-2}$ for J, U, and μ_{diff}, respectively, translating into a relative error of $<0.1\%$ ($<0.15\%$) for J (U). When the number of snapshots is decreased, the error begins to slowly increase. The experimentally least demanding case of 1000 snapshots per experimental realization results in an averaged estimated relative error of $<0.15\%$ ($<0.25\%$) for J (U). Although the accuracy of the estimation increases with decreasing statistical error, the low absolute estimation error for 2500 measurement snapshots may be explained with the excellent resilience neural-network approaches typically show with respect to noise. The higher value of the absolute error for the on-site repulsion and the chemical-potential differences can be explained with the known experimental precision. In particular, the absolute experimentally known precision interval is here set to be 0.02 for both the on-site repulsions as well as the chemical-potential differences, whereas it is of half the size for the hopping parameters. When rescaling the absolute errors through division with the experimental precision, the obtained errors (relative improvements of the known precision) are of similar magnitude for all parameters, see Appendix E.

A second-order error on the computed expectation values is given by faulty prepared initial states. The initial state is prepared with an accuracy of ≈99% per atom [14]. We modeled this imprecision with a probability p, that a single atom is not placed on the dedicated site but at one of its neighboring sites. An accuracy of 99% per atom thus corresponds to $p = 1\%$. Figure 3 demonstrates the dependence of

the absolute reconstruction error $|\Delta X|$ on the probability p for 2500 measurement snapshots. As the probability to prepare a faulty initial state increases, due to the flexibility of the neural-network approach only a small increase in the reconstruction error $|\Delta X|$ is observed. For the experimentally expected value $p \sim 1\%$, we are therefore still able to reduce the initial experimental error ($\pm 0.5\%$ around the average values $J \approx 1, U \approx 2$, $\mu \approx 1$) by half an order of magnitude.

IV. BAYESIAN ESTIMATION

To assess the quality of the results provided by the neural-network algorithm explained above, we aim to establish a reliable benchmark. In particular, we make a comparison with Bayesian parameter estimation. We note that a Bayesian classifier is provably optimal for classification problems [60] and we expect similarly that Bayesian parameter estimation will make optimal use of the measurement data. However, as we explain below, as a consequence of the size of the parameter space, we cannot use the full potential of the Bayesian estimation due to the arising computational challenges. We can, however, still perform a Bayesian estimate for a slightly less complex problem in order to have a guiding threshold for the size of the estimation error.

Bayesian inference can be used to obtain the set of most likely parameters θ associated with measurement outcomes M. More specifically, this goal is achieved by connecting the probability distribution over the parameter space Σ with the experimental observations via Bayes' theorem. In particular, for a measurement outcome M and a parameter $\theta \in \Sigma$ we have

$$P(\theta \mid M) = \frac{P(M \mid \theta)P(\theta)}{P(M)}, \tag{3}$$

where the likelihood $P(M \mid \theta)$ is the conditional probability to observe M given the underlying parameter θ, while $P(\theta \mid M)$ corresponds to the probability that the underlying parameter is θ given the measurement outcome M. $P(\theta)$ is the prior knowledge about the probability distribution of θ and $P(M)$ is the probability for M and serves as normalization constant. Note that $P(M \mid \theta)$ can be obtained by analyzing the statistics of the experimental outcomes for a range of parameters θ, and Bayes' rule (3) allows us to determine the most probable θ given experimental outcomes. Let us now discuss how this applies to our example.

The key simplification we have to make in comparison to the neural network is the discretization of the parameter space. When we train the machine-learning model, we pick values of the Hamiltonian parameters randomly during the training from the desired confidence interval and the model then interpolates over the continuous interval. Bayes' rule, on the other hand, allows us to calculate likelihood for a set of candidate parameter values and choose the one that best fits with the measurement data. To find which parameter combination results in maximum likelihood, we need to evaluate the model on a finite grid.

We restrict ourselves to the estimation of hopping amplitudes and on-site interaction terms. On the basic building block subset of 2×4 sites, this yields 18 parameters to estimate. If we were to select the optimal values from only five candidates per parameter, the number of parameter combinations to evaluate would be $5^{18} \approx 10^{12}$. As a consequence, we would need to perform an intractable number of these simulations to find which specific parameter combination best fits to the experimental observation.

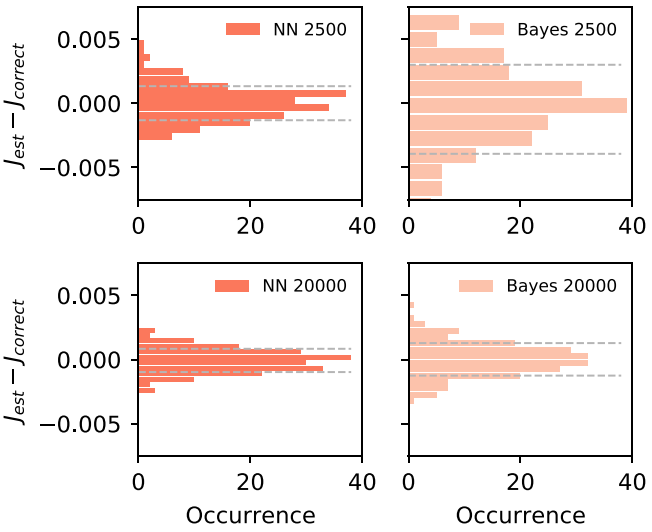


FIG. 4. Comparison of the performance of the neural network (left) and Bayesian method (right). The occurrences of the difference between estimated hopping and correct value are plotted as histogram for a set of 20 data sets each containing 2500 measurement snapshots (upper two panels) and 20 data sets each containing 20 000 measurement snapshots (lower two panels). The standard deviation of the error is indicated via gray dashed lines.

Since we cannot test every point of the large parameter space probability distribution in a reasonable computational time, we factorize this distribution as detailed in Appendix D. Essentially, we only estimate a small number of parameters at once, keep the rest fixed at the mean values of confidence interval and iterate until all estimated parameters have converged. The factorization reduces the number of quench simulations we need to perform to around 10^5. We consider a grid of 13 (21) candidates for each J (U) and estimate all parameters with relative error $<0.25\%$ for the experimentally most challenging case of 20 000 snapshots. This number could be further explored by using a finer candidate grid. However, this estimate comes already at a very high computational cost: When implemented with highly parallelized code [61], each instance of the Bayesian estimation takes approximately 24 hours per measurement set when running the algorithm on 36 cores. Additionally, there is no training and evaluation phase, the whole algorithm is rerun for each obtained measurement set—further adding to the computational demands over time.

We compare the performance of the neural-network-based method with the results obtained via the Bayesian inference benchmark described here, see Fig. 4. We show the distribution of estimation errors over 20 measurement sets for both the neural network and Bayesian estimation for 2500 and 20 000 measurement snapshots. We observe that the machine-learning algorithm shows a narrower error distribution and is, thus, firmly in the regime of errors comparable to Bayesian

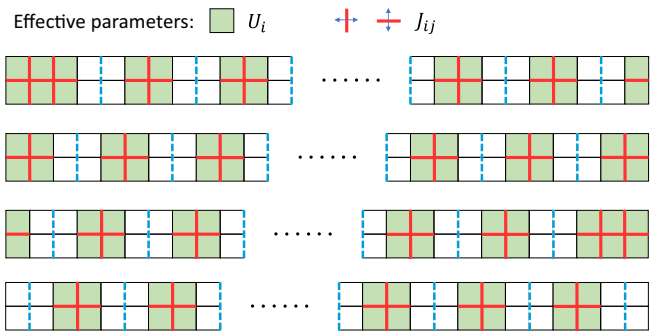


FIG. 5. Scaling of the protocol for the lattices of an arbitrary size. The black squares indicate the lattice with the dashed blue edge indicating a suppressed hopping term that defines a 2×4 subsystem. The shaded boxes and the thick red edges correspond to the repulsion and hopping terms, respectively.

estimation while requiring only a fraction of the computational cost once trained. Additionally, we observe that, with increasing number of measurement snapshots, Bayesian estimation approaches the neural-network error distribution. The neural-network estimator does appear to suffer significantly less from error distribution broadening when decreasing number of measurements, thus outperforming the Bayesian estimation benchmark.

All code needed to recreate these results as well as a minimal Hamiltonian reconstruction demonstration can be found in Ref. [61].

V. SCALING

In the previous sections we have shown that we are able to estimate parameters with relative error $\approx$0.1% ($\approx$0.15%) for J (U) from as little as 2500 measurement snapshots on a small subsystem of four bosons in eight lattice sites of a bosonic ladder lattice. The question remains how to experimentally scale the method up to larger systems while keeping the computational cost low. We first want to stress that the key ingredient of our scaling scheme, creating subsystems, is experimentally practical. Optical potentials can be projected with high resolution using holographic beam shaping [50]. Despite the influence on system-site potentials being in principle negligible due to the high-precision state-of-the-art optical potentials reach, to achieve a thorough validation we additionally account for the possibility of potential offsets on the subsystems boundaries induced by the projected optical potentials. For the following analysis, we therefore assume the parameters close to the introduced boundaries to be affected by the projected potential. Using the site numbering introduced in Fig. 1, this concerns the hopping amplitude between sites 1 and 5 (as well as 4 and 8) and the effective on-site interaction in these sites. We therefore develop a strategy of "shifting" our subsystem window such that we can take advantage of subsystem structure while capturing the whole system using as few measurements as possible.

We show this strategy in Fig. 5. Each row corresponds to one realization of the experiment, the blue lines denote where the boundary should be risen and green squares and red lines correspond to interactions and hopping amplitudes respectively we want to infer. Our strategy is then the following: Step 1: raise boundaries in the system. Step 2: Take a projective measurement of the whole system and repeat for the desired number of snapshots. Step 3: Use the relevant part of the data to calculate expectation values for each subsystem prepared this way. Step 4: Shift the boundary one site to the left and repeat steps 2 through 4. As illustrated in Fig. 5, we only need to repeat this process four times since the fifth shift would lead to a system configuration that we have already measured. In practice this means that for our ladder example, we can cover a ladder geometry lattice of arbitrary size with $4\times N$ snapshots. Note that overlapping of the boxes ensure that we can propagate the reference site for the evaluation of the chemical potential throughout the system. The final number of required measurements depends on the desired precision as decided according to the information provided in Fig. 3. For instance, considering 2500 snapshots per boundary configuration we find that we can fully learn the Hamiltonian of the $2\times M$ ladder lattice for any size of M with 10 000 experimental measurements.

Our algorithm and the scaling scheme are applicable in quantum gas experiments. We follow the experimental setup of Refs. [13,14] with respect to the choice of possible measurements. The preparation of the initial state can be prepared with about 99% per atom in state-of-the-art experiments by isolating particles from a Mott insulator [14]. In addition, optimal measurement fidelity is a result of the choice of the $2\times N$ system. More specifically, the detection fidelity in quantum gas experiments of 99% can be significantly improved by postselecting measurements on the initial atom number. Measurement errors can therefore be neglected in comparison with the statistical error resulting from only measuring a finite number of snapshots [14]. Cold atom experiments offer highly controlled systems and exceptional isolation from the environment. The coupling between subsystems is also negligible. The next-nearest-neighbor hopping J' can be suppressed by increasing lattice depth with respect to the nearest-neighbor coupling J. A ratio of $J'/J \approx 10^{-3}$ is achieved already at a depth of $V_r \approx 20E_r$, where E_r corresponds to the recoil energy. More shallow lattices can be addressed by including the next-nearest neighbor coupling in the Hamiltonian. In leading order, this can be achieved by adding a single additional parameter J' which can be assumed to be constant across the system. We believe these facts make our scheme a potential approach for the quantum simulation verification.

VI. DISCUSSION AND CONCLUSIONS

Precise and robust Hamiltonian learning techniques that work reliably using only experimental data are key for the verification of quantum simulators and the manifestation of quantum advantage. The work presented here opens an avenue towards achieving just that and is immediately applicable to state-of-art quantum simulation experiments. In particular, we have developed a feed-forward neural-network-based approach towards out-of-equilibrium Hamiltonian learning of large-scale quantum simulators. By making use of the local structure of simulated Hamiltonians we build up a subsystem-based Hamiltonian learning method. We use Bayesian estimation on a discrete parameter space as an estimation benchmark. We show that the neural-network

approach outperforms the guiding Bayesian learning benchmark at a fraction of the computational cost. We have illustrated the effectiveness of our method on the Hamiltonian reconstruction of an out-of-equilibrium bosonic system on a ladder in an optical lattice. Our specific example concerned the situation where the parameters can be relatively precisely estimated from first principles. Our method is then able to determine the parameters with a significantly higher precision than the experimentally accessible level. We show further examples of the usage of our algorithm in Appendix E.

When adapting our method to other systems the key challenge is the subsystem design. In the particular application explained here, we were able to select the subsystem small enough for us to be able to simulate the dynamics of the subsystem exactly. To increase the number of particles and fully capture a larger subsystem size, we may use a suitable approximate method to simulate the quantum dynamics [42,62,63] while approaching the factorization of the parameter space in an analogous fashion as described in this work.

ACKNOWLEDGMENTS

We are grateful for financial support from the Swiss National Science Foundation, the NCCR QSIT. This work has received funding from the European Research Council under Grant No. 771503.

APPENDIX A: EXPERIMENTAL IMPERFECTION SOURCES

We detail here the influence of sources of imperfection and decoherence beyond the parameter offsets modeled in the Bose-Hubbard Hamiltonian (1).

Experiments tailoring ultracold atoms in an optical lattice offer a high level of control and exceptional isolation from the environment. The main imperfection source is modeled by Hamiltonian (1). The next dominant term are next-nearest-neighbor couplings J'. Increasing lattice depth suppresses next-nearest-neighbor coupling with respect to the nearest-neighbor coupling J. A ratio of $J'/J \approx 10^{-3}$ is achieved already at a depth of $V_r \approx 20E_r$, where E_r corresponds to the recoil energy. More shallow lattices can be addressed by including the next-nearest-neighbor coupling in the Hamiltonian. In leading order, this can be achieved by adding a single additional parameter J' which can be assumed to be constant across the system.

Additional error sources can be a result of decoherence. Decoherence is mainly caused by quantum state collapse from photon scattering at the optical potentials and atom lifetime, which is limited by the vacuum pressure. We reasonably neglect the decoherence, as each decoherence time exceeds 10^3 tunneling times and we consider a time evolution of the order of a few tunneling times.

We therefore believe our method is not threatened by experimental imperfections.

APPENDIX B: MEASUREMENT SEQUENCE AND SIMULATIONS

In the following, we provide detailed information about the considered measurement sequence and its simulation. As explained in the main text, the goal is to provide an estimation of the parameters of the Bose-Hubbard Hamiltonian Eq. (1) based on atom density measurements after a time evolution of a prepared initial state. In particular, we choose an initial state with four atoms, which constitutes the upper limit of bosons on a 2×4 lattice that is still feasible for our approach because it is fast to simulate via exact diagonalization. We have found that the particular initial distribution of the four bosons on the lattice plays a negligible role in the precision of the parameter estimation and can therefore be chosen arbitrarily, or as a distribution which is experimentally most feasible. The initial state used exemplary here contains an atom on every second site. Measurements are taken after a time evolution under the Bose-Hubbard Hamiltonian (1) of T, where we here choose as an example $T = 200\hbar$ to be experimentally realizable. Testing with smaller T did not show any significant difference in precision, therefore the neural network may be trained for a specific experiment by using the most experimentally suitable time evolution T.

We follow the experimental setup of Refs. [13,14] with respect to the choice of possible measurements. In particular, the measurements considered here on the time-evolved state correspond to spatially resolved atom-number measurements. Experimentally, such a measurement in a cold atom setup can be performed by quickly ramping up the longitudinal lattice potential to the Mott insulating phase and thereby freezing the dynamics. The number of atoms per site can be determined by using fluorescence imaging after letting the atoms expand in transversal tubes to avoid parity projections [13,14]. We note here that a different experimental platform for quantum simulation requires a different experimental sequence.

We simulate the time evolution and projective measurements using exact diagonalization of the Bose-Hubbard Hamiltonian (1). In particular, the measurement snapshots correspond to projections of the time-evolved state on the Fock state basis. Numerically, the projection is straightforwardly taken by calculating the overlap between each Fock state $|i\rangle$ and the time-evolved state $|\Psi(T)\rangle$: The probability to measure $|i\rangle$ (corresponding to an atom number distribution s) is given by $|\langle i|\Psi(T)\rangle|^2$. Repeating this simulated projection N times on the state $|\Psi(T)\rangle$ yields a collection of snapshots $(s_1, \ldots, s_N)$.

We calculate density correlation expectation values directly from the snapshots, i.e.,

$$\langle n_1..n_l\rangle = \sum_{i=1}^{N} s_i(1)\cdots s_i(l), \tag{B1}$$

where $s_i(j)$ returns the number of atoms on the jth site of snapshot s_i. As a consequence, a statistical error as observed in measurements is introduced into the simulation of the expectation values.

APPENDIX C: NEURAL NETWORKS

In the following, we provide a detailed description of the neural-network-based Hamiltonian reconstruction scheme. For each of the hopping amplitudes J_{ij} and on-site repulsions U_i, a separate network is trained to estimate the parameter's

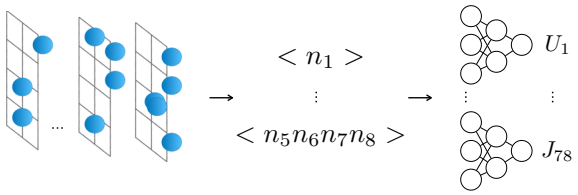


FIG. 6. Workflow of the neural-network-based algorithm: 1. Measurement snapshots are experimentally taken or numerically simulated. 2. Compute the set of relevant correlators. 3. The set of correlators then serves as an input of 25 feed-forward supervised networks, one for each parameter to be estimated.

value. We note here that estimating a set of parameters together lowers the precision of the parameter estimation. Separating the estimation into separate networks for each parameter does not correspond to a factorization of the output probability but is rather interpreted as a one-to-one mapping between a set of measurement outcomes and parameter configurations. Each network takes as input postprocessed measurement data and outputs the respective parameter value. In particular, from N density snapshots conducted on the 2×4 lattice in a first step, a set of expectation values are calculated in order to compress the data. More concretely, we calculate the following expectation values:

(i) the single-site densities $\langle n_i \rangle$ with n_i being the number of atoms on site i;

(ii) the two-point density correlation functions $\langle n_i n_j \rangle$, $i < j$;

(iii) the three-point density correlation functions $\langle n_i n_j n_k \rangle$, $i < j < k$;

(iv) the four-point density correlation functions $\langle n_i n_j n_k n_l \rangle$, $i < j < k < l$;

(v) $\langle n_i(n_i - 1) \rangle$ counting multiple occupancies on site i.

For eight lattice sites, this yields a total of 171 expectation values. This number, and consequently the network input dimension, is independent of the number of conducted measurement snapshots. For each parameter to be estimated, a network is trained taking as input 171 expectation values calculated from the experimental measurements to output the respective parameter's value. The parameter estimation workflow is depicted in Fig. 6.

We use a fully connected neural network with one input layer with 171 neurons (one neuron for each expectation value), five hidden layers with 300, 400, 300, 150, and 100 neurons, and an output layer with one neuron. The value of the output neuron corresponds to the parameter to be estimated. The network is trained using supervised learning and the mean squared loss function $C(X)$ for continuous parameter estimation is

$$C(X) = \frac{|X^{\text{pred}} - X^{\text{label}}|^2}{N_{\text{Batch}}}. \tag{C1}$$

Here, we use X as placeholder for the parameter J_{ij}, U_i, or $\mu_{\text{diff},i}$ to be estimated. X^{label} corresponds to the correct value whereas X^{pred} is the neural-network prediction. The batch size during training is given by N_{Batch}.

We create a training set of 150 500 training examples, each example consisting of the expectation values together with the respective parameter label. The time evolution and conducted measurements are simulated via exact diagonalization. For each of the training examples, the correct parameter values are chosen randomly within the interval of confidence, here $J_{ij} = 1 \pm 0.005$, $U_i = 2 \pm 0.01$, and $\mu_i = 1 \pm 0.005$. Using the learning rate $\eta = 10^{-5}$ and a batch size of $N_{\text{Batch}} = 50$, the training is conducted for a total of 2.1×10^6 training steps. The training and evaluation loss for the two networks estimating the parameters J_{23} and U_2 using 2500 measurement snapshots are shown in Fig. 7, where the evaluation set corresponds to a set of 500 examples not used for training. The neural network was trained using the library Tensorflow [64].

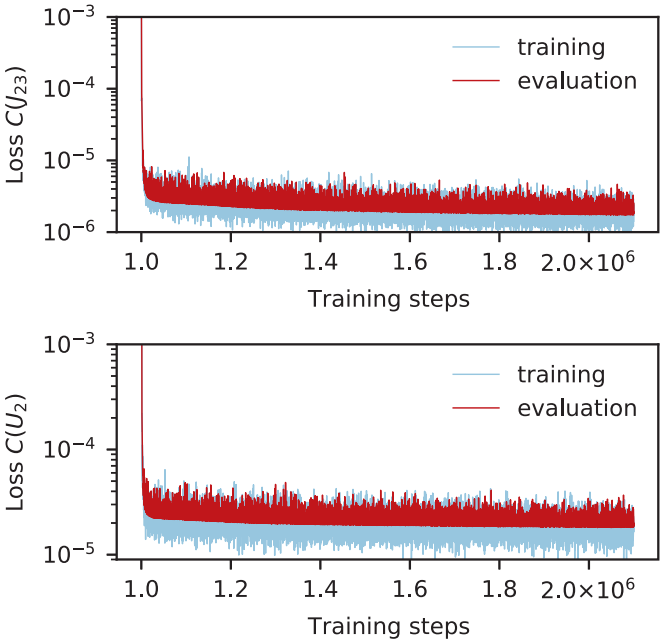


FIG. 7. Network training (light blue) and evaluation loss (dark red) as a function of training steps for the network estimating the hopping parameter J_{23} (upper panel) and the network estimating the on-site repulsion U_2 (lower panel).

The accuracy of the neural-network prediction can be quantified by the mean difference of the estimated values to the correct parameter values used in the simulation,

$$\Delta X = |X^{\text{pred}} - X^{\text{correct}}|, \tag{C2}$$

where $X \in J_{ij}, U_i, \mu_i$. We note the difference to the squared loss function (C1). To determine the correlation between estimation errors of different parameters within a measurement sequence, we plot the estimation error $\Delta J/J$ vs $\Delta U/U$ in Fig. 8, where J corresponds to the mean of all ten hopping amplitudes, and U to the mean of all eight on-site repulsions. Each point in Fig. 8 represents one experimental measurement sequence consisting of 2500 measurement snapshots. If the parameter estimation errors were correlated, we would expect the outcome to be distributed along the diagonal (or antidiagonal for anticorrelated errors). Figure 8 shows no such behavior. Thus we conclude that the hopping amplitudes and on-site repulsions can be estimated with independent precision.

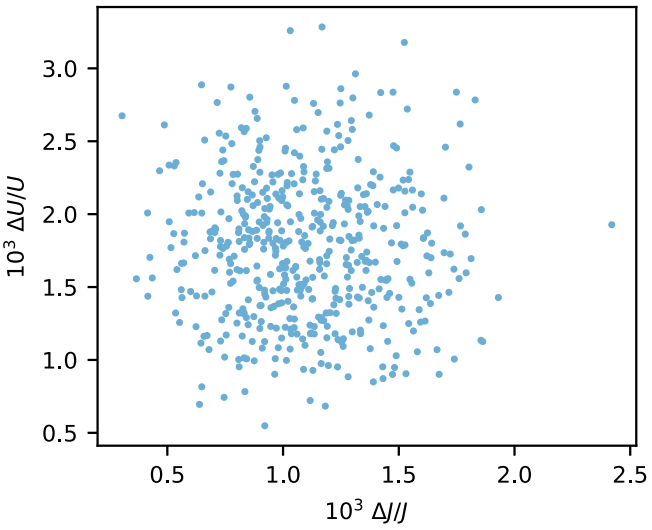


FIG. 8. The relative parameter estimation error $\Delta J/J$ vs $\Delta U/U$ for 2500 measurement snapshots.

APPENDIX D: BAYESIAN ESTIMATION

Bayesian inference is a statistical inference method for the calculation of the probability for a certain hypothesis based on available information [65–67]. Specifically, let us consider a machine, where a hypothesis corresponds to a parameter configuration θ. Different machines specified by different θ also differ in the probability distribution of their measurable outcome S. The goal is to find which hypothesis (parameter configuration θ) is most probable given the measurement S. The relation of the internal parameters θ and accessible observables S is specified by Bayes' rule:

$$P(\theta \mid S) = \frac{P(S \mid \theta)P(\theta)}{P(S)}, \tag{D1}$$

where

(i) θ can be any set of parameters specifying the machine. Our task is to determine which θ is most probable according to the measured data S.

(ii) $P(\theta)$ is the so-called prior probability. It contains our initial knowledge about the probability distribution of the parameters θ before taking into account the data S.

(iii) $P(\theta \mid S)$ is the posterior probability specifying the conditional probability of a specific parameter configuration θ given the measured data S.

(iv) $P(S \mid \theta)$ is the so-called likelihood: the probability of observing S given θ.

(v) $P(S) = \sum_\theta P(S \mid \theta)P(\theta)$ is the normalization factor.

In our specific case the parameters θ correspond to all the unknown hopping amplitudes and on-site interactions $\theta = \{\vec{J}, \vec{U}\}$, and S corresponds to quantum gas microscope measurement M (snapshots).

Note that in our case we do not only have a single observation S but many thousands of them. For the case of many observations Bayes' rule generalizes as follows: For $\{S\} = (s_1, s_2, \ldots, s_n)$, where each observation s_i is independent and identically distributed (i.i.d.), we can formulate the Bayes' theorem as [68,69]

$$P(\theta \mid \{S\}) = \frac{P(\{S\} \mid \theta)P(\theta)}{P(\{S\})}, \tag{D2}$$

where the combined likelihood of a set of observations is given by the product of likelihoods for each individual observation,

$$P(\{S\} \mid \theta) = \prod_i P(s_i \mid \theta). \tag{D3}$$

In our case the observations $\{s_i\}$ correspond to all the possible Fock states one can obtain by occupational projections. The likelihood function $P(\{S\}|\theta)$ is calculated according to Eq. (D3), given by the product of probabilities for measuring the corresponding Fock states:

$$P(\{S\} \mid \theta) = \prod_i P(s_i \mid \theta) = \prod_i P(|i\rangle \mid \theta). \tag{D4}$$

As for the prior, we consider a uniform distribution over the all the possible parameter configurations

$$P(\theta) = \frac{1}{n_p}, \tag{D5}$$

where n_p is the number of candidate parameter configurations.

Consider the system shown in Fig. 1, we have $n_p \approx 10^{18}$ when we assign ten candidates for each J and U. This parameter space size exceeds the capacity of any contemporary computer. Therefore we need to further factorize the parameter estimation process. We estimate a small group of parameters each time and update the values, iterating over all parameter groups. We begin with ten hopping amplitudes J, then move on to eight on-site repulsions U and consequently we come back to re-estimate J and U. The hopping amplitudes J and on-site repulsions U are in addition factorized in smaller subgroups accordingly to their spatial location. The details of this factorization process are shown in Fig. 9. We prepare eight groups of U and three groups of J.

We discretize the parameter space accordingly to the experimentally known uncertainty. In particular, for the experimental precision of $\pm 0.5\%$, the correct values of the parameters lie within the intervals [0.995,1.005] for J, [1.99,2.01] for U, and [0.995,1.005] for μ and we choose 13 parameter candidates for J and 21 candidates for U as equally spaced within these intervals. During the estimation process, the simulations are performed by using the mean value of the chemical potential within the known precision, i.e., $\mu = 1$. As the correct value may differ, an error is introduced by this choice. The accuracy of the estimation process might increase by including the estimation of the chemical potential, we here restrict ourselves to the estimation of J and U due to the computational cost.

The iterative parameter estimation is performed as follows: The estimation is based on N experimental snapshots which are re-used in every iteration. As we factorized the parameter estimation, we only estimate the parameters within a subgroup at a time, while keeping the rest of the parameters fixed. In particular, we initially set the values of the parameters outside of the current subgroup to the mean values within the experimental uncertainty (i.e., $J = 1.0$, $U = 2.0$) and update them accordingly to their estimated values after each estimation.

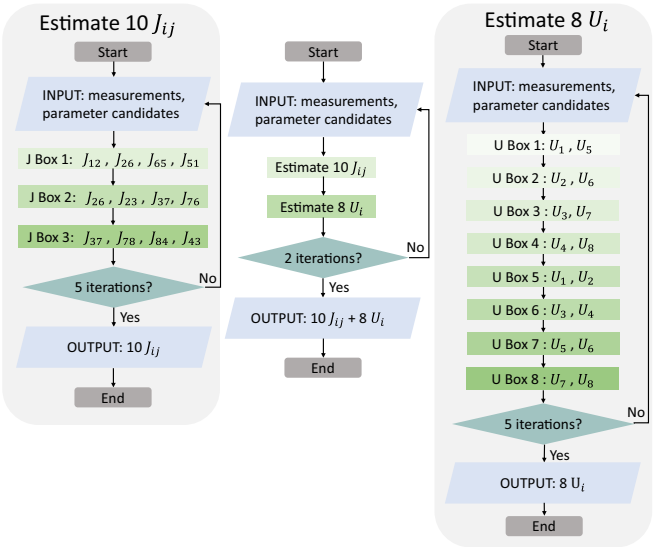


FIG. 9. Workflow of the Bayesian algorithm: due to the large parameter space, we evaluate in groups while keeping the rest constant according to the grouping scheme shown in the flowchart. One iteration corresponds to evaluating all three (eight) subgroups of J (U).

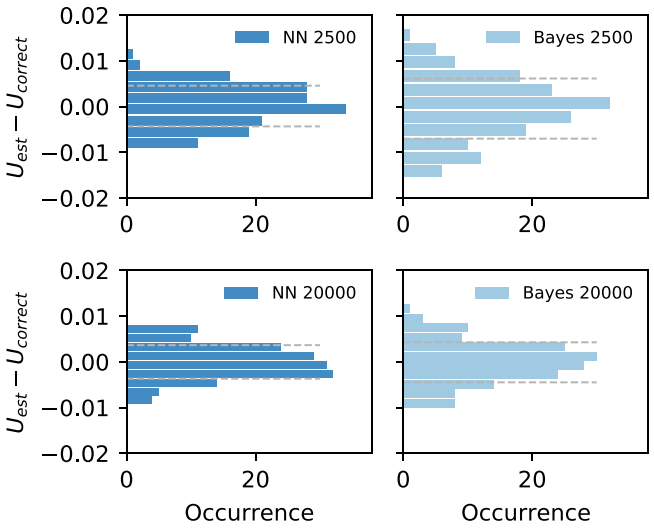


FIG. 10. Comparison of the performance of the neural network (left) and Bayes' method (right). The occurrences of the difference between estimated on-site repulsion and correct value are plotted as histogram for a set of 20 data sets each containing 2500 measurement snapshots (upper two panels) and 20 data sets each containing 20 000 measurement snapshots (lower two panels). The standard deviation of the error is indicated via gray dashed lines.

We start the parameter estimation process by estimating the hopping amplitudes. Concretely, we subsequently estimate the J_{ij} per subgroup (three subgroups) and update the parameter values accordingly. Within each subgroup of the hopping amplitudes, we estimate four hopping amplitudes at a time. Using 13 parameter candidates for each hopping amplitude, this yields a total dimension of $13^4 = 28\,561$ for the parameter space. After the first estimation of all hopping amplitudes, the on-site repulsions are estimated by subgroup. As each subgroup here only contains two U_i to be estimated, we can choose a finer grid (21 candidates), while keeping a feasible parameter space dimension of $21^2 = 441$. Estimating all on-site repulsions completes the first iteration, and the process is repeated. In total, we perform five iterations of the estimation process.

In summary of this Appendix section, we employ Bayes' rule in the following steps:

(1) Prepare "experimental" data: We initialize the system in a selected Fock state, simulate unitary evolution under a Hamiltonian with randomly selected correct parameters, followed by a projective measurement. We repeat this process by the number of times that corresponds to the selected number of snapshots, N. We only keep the measurements $\{S\}$ for the follow-up estimation.

(2) Simulated likelihood: We simulate the unitary evolution under each Hamiltonian specified by all the possible candidate configurations from the parameter space. We factorize this process as shown in Fig. 9. For each candidate parameter configuration θ, we take the final state (after time evolution T) and calculate the overlap with all the Fock basis states respectively: $\{\langle\Psi_{\text{final},\theta}|i\rangle\}$. This yields the ingredients $\{P(|i\rangle \mid \theta)\}$ we need to calculate the likelihood function as shown in Eq. (D4).

(3) Select the most likely parameters: The parameter candidates θ for which the set $\{P(|i\rangle \mid \theta)\}$ can maximize the likelihood (D4) is most likely to generate the same physics observed in the experiment.

(4) Update: Update the current values of the parameters in the current group by the Bayesian estimated values.

We have shown the comparison between the neural-network based estimation and the Bayesian inference in Fig. 4 for the estimation of hopping amplitudes, where the neural network outperforms the Bayesian estimation. We show in addition the results for the comparison between the estimation of the on-site repulsions provided by the neural network and using Bayesian inference in Fig. 10. While the neural network still yields more accurate results, the difference is less striking compared with the hopping amplitude estimation in Fig. 4.

APPENDIX E: ADDITIONAL PARAMETER REGIMES

In the following, we examine the dependence of the parameter estimation accuracy for varying parameter regimes. We start by analyzing the dependence on the size of the parameter confidence interval, i.e., the previously known experimental precision. In the main text, all simulations are done with a previously known parameter precision of 1% (±0.5%) and the neural-network estimation improves this precision by half an order of magnitude.

The initially known precision might vary from this choice. In particular, we consider parameter precisions up to ±2.5% and examine the relative improvement in precision obtained after applying the neural-network Hamiltonian reconstruction scheme. For each interval, we retrain neural networks with training data within the experimentally known precision. The relative improvement of the initial precision P, i.e., the ratio $\Delta X/P$ is shown in Fig. 11 for $X = J, U$. Using

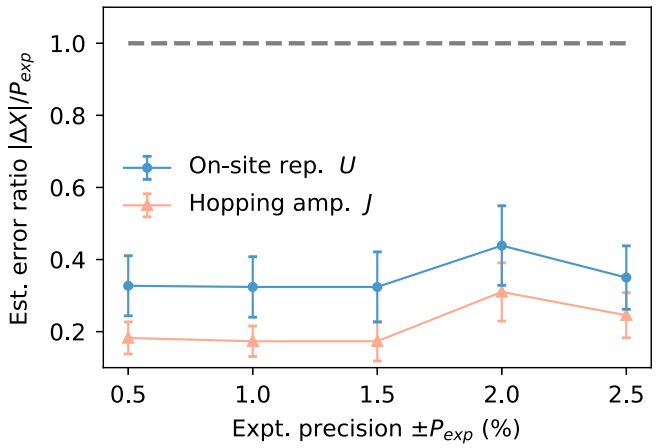


FIG. 11. The ratio of the parameter estimation error of the on-site repulsion $\Delta X = \Delta U$ and the hopping strengths $\Delta X = \Delta J$ divided by the experimental precision P_{expt} as a function of P_{expt}. The gray dashed line indicates the ratio $P_{\text{expt}}/P_{\text{expt}} = 1$ to emphasize the improvement with respect to the experimental precision. We use here the neural-network-based parameter estimation with 5000 snapshots.

5000 snapshots, we obtain a significant gain in precision which does not notably increase within the interval $P \in [0.5\%, 2.5\%]$.

We examine in addition how the parameter estimation precision varies with respect to the ratio of the on-site repulsion U to the hopping strength J. In particular, we vary the (average) strength $U_{\text{mean}} = 0.5, \ldots, 70$ of the on-site repulsion $U_i = U_{\text{mean}} \pm 0.01$ while keeping the hopping amplitudes fixed ($J_{\text{mean}} = 1.0$) in the interval $J_{ij} = J_{\text{mean}} \pm 0.005$. The *absolute* precision of U is unchanged for varying U_{mean} in order to ensure a meaningful comparison of the obtained parameter estimation precisions. The absolute parameter estimation errors

$$\Delta U = \frac{1}{8} \sum_i \left| U_i^{\text{pred}} - U_i^{\text{label}} \right|, \tag{E1}$$

$$\Delta J = \frac{1}{10} \sum_{\langle ij \rangle} \left| J_{ij}^{pred} - J_{ij}^{label} \right| \tag{E2}$$

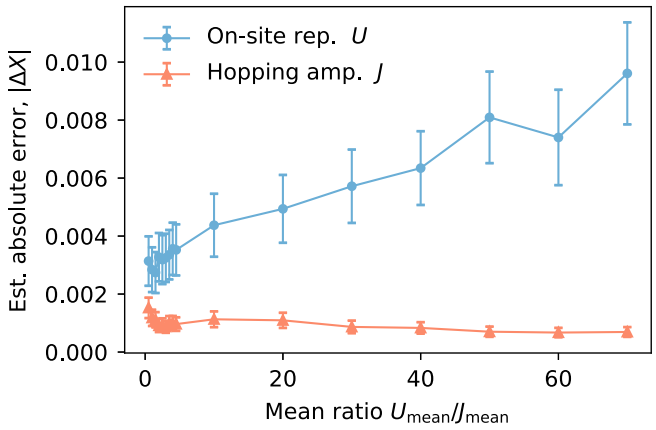


FIG. 12. The absolute parameter estimation errors of the on-site repulsion $\Delta X = \Delta U$ and the hopping strengths $\Delta X = \Delta J$ as a function of the ratio $U_{\text{mean}}/J_{\text{mean}}$. We use here the neural-network based parameter estimation with 5000 snapshots.

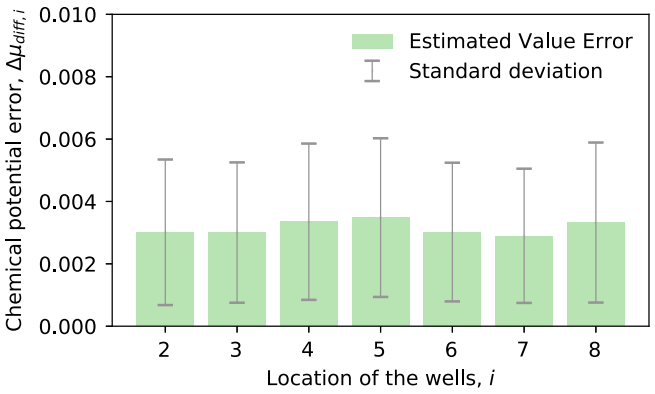


FIG. 13. Estimation error of the neural-network estimation scheme for the chemical-potential differences $\mu_{i,\text{diff}} = \mu_i - \mu_1$ for 2500 measurement snapshots as a function of the spatial location i, averaged over 500 data sets.

are plotted in Fig. 12 as a function of $U_{\text{mean}}/J_{\text{mean}}$ using 5000 snapshots. Here, the sum in the second line runs over all nearest-neighbor bonds.

Figure 12 shows an increasing absolute error in estimating the value of the on-site repulsion for increasing ratio $U_{\text{mean}}/J_{\text{mean}}$. At the same time, the estimation error of the hopping amplitudes slightly decreases for increasing $U_{\text{mean}}/J_{\text{mean}}$. We can understand this result by examining the interplay of the on-site repulsion and the hopping strengths. More specifically, the timescale of the atom movement between the different sites is set by the hopping strength J_{mean}. The on-site repulsion might be roughly understood as influencing the "direction" of hopping: The stronger is the on-site repulsion, the more the probability of measuring a configuration with several atoms on a site is reduced. As our simulations take place with four atoms on eight lattice sites, the configuration is sufficiently sparse such that double occupancies may be avoided. For very large ratio $U_{\text{mean}}/J_{\text{mean}}$, the probability to measure a double occupancy becomes increasingly small and therefore unlikely to be observed with a finite number of measurement snapshots. Small changes in the on-site repulsion strength are therefore expected to be increasingly difficult to

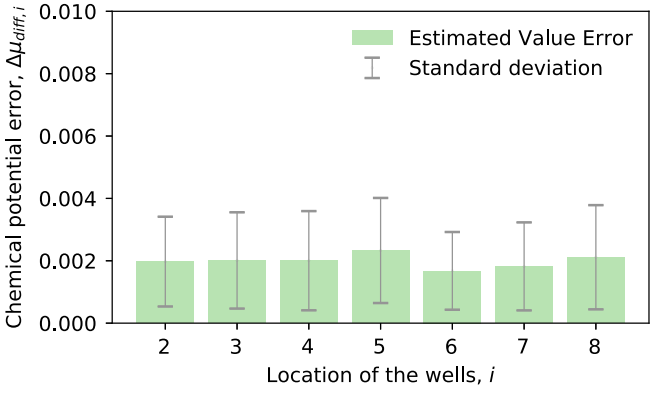


FIG. 14. Estimation error of the neural-network estimation scheme for the chemical-potential differences $\mu_{i,\text{diff}} = \mu_i - \mu_1$ for 20 000 measurement snapshots as a function of the spatial location i, averaged over 500 data sets.

detect because the effect of the on-site repulsion manifests itself mainly in the occurrence of multiple occupancies of lattice sites. This behavior is shown in Fig. 12. A more dense configuration of atoms (i.e., a more than half-filled lattice) might show a different behavior, as the probability of multiple occupancies is not suppressed in the same way compared with a sparse atom configuration.

APPENDIX F: ESTIMATING CHEMICAL POTENTIALS: DETAILS

As we are considering only states with a constant total atom number (in particular, we use $N = 4$ atoms), a uniform chemical potential μ yields only a global phase factor $e^{-i/\hbar\mu N}$. Thus, a uniform chemical potential does not affect the system's dynamics and is as a consequence not detectable within the closed system. Relevant and measurable effects are instead induced by a *nonuniform* chemical potential. We can quantify the deviations of the chemical potential by considering, e.g., the differences $\mu_{\mathrm{diff},i} := \mu_i - \mu_1$, $i \geqslant 2$.

When estimating the parameters $\{J_{ij}, U_i, \mu_{\mathrm{diff},i}\}$, we vary all parameters within a percent, i.e., $\mu = 1 \pm 0.005$ (and $J = 1.0 \pm 0.005$, $U = 2.0 \pm 0.01$). As a consequence, the differences $\mu_{\mathrm{diff},i} = \mu_i - \mu_1$ take the values $\mu_{\mathrm{diff},i} = 0 \pm 0.01$. We show the obtained estimation error

$$\Delta\mu_{\mathrm{diff},i} = \left|\mu^{\mathrm{label}}_{\mathrm{diff},i} - \mu^{\mathrm{est.}}_{\mathrm{diff},i}\right| \tag{F1}$$

and their spatial dependence in Figs. 13 and 14 for 2500 and 20 000 measurement snapshots, respectively.

[1] R. Blatt and C. F. Roos, Quantum simulations with trapped ions, Nat. Phys. **8**, 277 (2012).

[2] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller *et al.*, Real-time dynamics of lattice gauge theories with a few-qubit quantum computer, Nature (London) **534**, 516 (2016).

[3] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).

[4] D. Zhu, S. Johri, N. Linke, K. A. Landsman, C. H. Alderete, N. H. Nguyen, A. Y. Matsuura, T. H. Hsieh, and C. Monroe, Generation of thermofield double states and critical ground states with a quantum computer, Proc. Natl. Acad. Sci. USA **117**, 25402 (2020).

[5] P. Roushan, C. Neill, A. Megrant, Yu. Chen, R. Babbush, R. Barends, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth *et al.*, Chiral ground-state currents of interacting photons in a synthetic magnetic field, Nat. Phys. **13**, 146 (2017).

[6] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare *et al.*, Observation of topological phenomena in a programmable lattice of 1,800 qubits, Nature (London) **560**, 456 (2018).

[7] R. Ma, B. Saxberg, C. Owens, N. Leung, Y. Lu, J. Simon, and D. I. Schuster, A dissipatively stabilized Mott insulator of photons, Nature (London) **566**, 51 (2019).

[8] B. Chiaro, C. Neill, A. Bohrdt, M. Filippone, F. Arute, K. Arya, R. Babbush, D. Bacon, J. Bardin, R. Barends, S. Boixo, D. Buell, B. Burkett, Y. Chen, Z. Chen, R. Collins, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen *et al.*, Direct measurement of non-local interactions in the many-body localized phase, arXiv:1910.06024.

[9] T. Hensgens, T. Fujita, L. Janssen, X. Li, C. J. Van Diepen, C. Reichl, W. Wegscheider, S. D. Sarma, and L. M. K. Vandersypen, Quantum simulation of a Fermi–Hubbard model using a semiconductor quantum dot array, Nature (London) **548**, 70 (2017).

[10] C. Gross and I. Bloch, Quantum simulations with ultracold atoms in optical lattices, Science **357**, 995 (2017).

[11] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner *et al.*, Probing many-body dynamics on a 51-atom quantum simulator, Nature (London) **551**, 579 (2017).

[12] F. Görg, M. Messer, K. Sandholzer, G. Jotzu, R. Desbuquois, and T. Esslinger, Enhancement and sign change of magnetic correlations in a driven quantum many-body system, Nature (London) **553**, 481 (2018).

[13] M. Rispoli, A. Lukin, R. Schittko, S. Kim, M. E. Tai, J. Léonard, and M. Greiner, Quantum critical behaviour at the many-body localization transition, Nature (London) **573**, 385 (2019).

[14] A. Lukin, M. Rispoli, R. Schittko, M. E. Tai, A. M. Kaufman, S. Choi, V. Khemani, J. Léonard, and M. Greiner, Probing entanglement in a many-body–localized system, Science **364**, 256 (2019).

[15] B. Yang, H. Sun, R. Ott, H.-Y. Wang, T. V. Zache, J. C. Halimeh, Z.-S. Yuan, P. Hauke, and J.-W. Pan, Observation of gauge invariance in a 71-site Bose–Hubbard quantum simulator, Nature (London) **587**, 392 (2020).

[16] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho *et al.*, Quantum phases of matter on a 256-atom programmable quantum simulator, Nature (London) **595**, 227 (2021).

[17] K. Wintersperger, C. Braun, F. N. Ünal, A. Eckardt, M. Di Liberto, N. Goldman, I. Bloch, and M. Aidelsburger, Realization of anomalous Floquet topological phases with ultracold atoms, Nat. Phys. **16**, 1058 (2020).

[18] J. R. Garrison and T. Grover, Does a Single Eigenstate Encode the Full Hamiltonian? Phys. Rev. X **8**, 021026 (2018).

[19] X.-L. Qi and D. Ranard, Determining a local Hamiltonian from a single eigenstate, Quantum **3**, 159 (2019).

[20] E. Chertkov and B. K. Clark, Computational Inverse Method for Constructing Spaces of Quantum Models from Wave Functions, Phys. Rev. X **8**, 031029 (2018).

[21] E. Bairey, I. Arad, and N. H. Lindner, Learning a Local Hamiltonian from Local Measurements, Phys. Rev. Lett. **122**, 020504 (2019).

[22] E. Bairey, C. Guo, D. Poletti, N. H. Lindner, and I. Arad, Learning the dynamics of open quantum systems from their steady states, New J. Phys. **22**, 032001 (2020).

[23] T. J. Evans, R. Harper, and S. T. Flammia, Scalable Bayesian hamiltonian learning, arXiv:1912.07636.

[24] C. Cao, S.-Y. Hou, N. Cao, and B. Zeng, Supervised learning in Hamiltonian reconstruction from local measurements on eigenstates, J. Phys.: Condens. Matter **33**, 064002 (2020).

[25] X. Turkeshi, T. Mendes-Santos, G. Giudici, and M. Dalmonte, Entanglement-Guided Search for Parent Hamiltonians, Phys. Rev. Lett. **122**, 150606 (2019).

[26] X. Turkeshi and M. Dalmonte, Parent Hamiltonian reconstruction of Jastrow-Gutzwiller wavefunctions, SciPost Phys. **8**, 042 (2020).

[27] L. Che, C. Wei, Y. Huang, D. Zhao, S. Xue, X. Nie, J. Li, D. Lu, and T. Xin, Learning quantum Hamiltonians from single-qubit measurements, Phys. Rev. Research **3**, 023246 (2021).

[28] S. G. Schirmer, A. Kolli, and D. K. L. Oi, Experimental Hamiltonian identification for controlled two-level systems, Phys. Rev. A **69**, 050306(R) (2004).

[29] A. Valenti, E. van Nieuwenburg, S. Huber, and E. Greplova, Hamiltonian learning for quantum error correction, Phys. Rev. Research **1**, 033092 (2019).

[30] E. Greplova, C. K. Andersen, and K. Mølmer, Quantum parameter estimation with a neural network, arXiv:1711.05238.

[31] S. J. Devitt, J. H. Cole, and L. C. L. Hollenberg, Scheme for direct measurement of a general two-qubit Hamiltonian, Phys. Rev. A **73**, 052317 (2006).

[32] J. H. Cole, S. G. Schirmer, A. D. Greentree, C. J. Wellard, D. K. L. Oi, and L. C. L. Hollenberg, Identifying an experimental two-state Hamiltonian to arbitrary accuracy, Phys. Rev. A **71**, 062312 (2005).

[33] J. H. Cole, S. J. Devitt, and L. C. L. Hollenberg, Precision characterization of two-qubit Hamiltonians via entanglement mapping, J. Phys. A: Math. Gen. **39**, 14649 (2006).

[34] D. Burgarth, K. Maruyama, and F. Nori, Coupling strength estimation for spin chains despite restricted access, Phys. Rev. A **79**, 020305(R) (2009).

[35] E. B. Av, Y. Shapira, N. Akerman, and R. Ozeri, Direct reconstruction of the quantum-master-equation dynamics of a trapped-ion qubit, Phys. Rev. A **101**, 062305 (2020).

[36] J. Eisert, M. Friesdorf, and C. Gogolin, Quantum many-body systems out of equilibrium, Nat. Phys. **11**, 124 (2015).

[37] Z. Li, L. Zou, and T. H. Hsieh, Hamiltonian Tomography Via Quantum Quench, Phys. Rev. Lett. **124**, 160502 (2020).

[38] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Hamiltonian Learning and Certification Using Quantum Resources, Phys. Rev. Lett. **112**, 190501 (2014).

[39] D. Hangleiter, I. Roth, J. Eisert, and P. Roushan, Precise hamiltonian identification of a superconducting quantum processor, arXiv:2108.08319.

[40] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[41] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[42] I. Cirac, D. Perez-Garcia, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, and theorems, Rev. Mod. Phys. **93**, 045003 (2021).

[43] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert, Architectures for Quantum Simulation Showing a Quantum Speedup, Phys. Rev. X **8**, 021010 (2018).

[44] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega, Closing Gaps of a Quantum Advantage with Short-Time Hamiltonian Dynamics, Phys. Rev. Lett. **125**, 250501 (2020).

[45] S. S. Kalantre, J. P. Zwolak, S. Ragole, X. Wu, N. M. Zimmerman, M. D. Stewart, and J. M. Taylor, Machine learning techniques for state recognition and auto-tuning in quantum dots, npj Quantum Inf. **5**, 6 (2019).

[46] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller, Cold Bosonic Atoms in Optical Lattices, Phys. Rev. Lett. **81**, 3108 (1998).

[47] I. Bloch, J. Dalibard, and W. Zwerger, Many-body physics with ultracold gases, Rev. Mod. Phys. **80**, 885 (2008).

[48] H. A. Gersch and G. C. Knollman, Quantum cell model for bosons, Phys. Rev. **129**, 959 (1963).

[49] M. Aidelsburger, M. Atala, M. Lohse, J. T. Barreiro, B. Paredes, and I. Bloch, Realization of the Hofstadter Hamiltonian with Ultracold Atoms in Optical Lattices, Phys. Rev. Lett. **111**, 185301 (2013).

[50] P. Zupancic, P. M. Preiss, R. Ma, A. Lukin, M. E. Tai, M. Rispoli, R. Islam, and M. Greiner, Ultra-precise holographic beam shaping for microscopic quantum control, Opt. Express **24**, 13881 (2016).

[51] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice, Nature (London) **462**, 74 (2009).

[52] J. F. Sherson, C. Weitenberg, M. Endres, M. Cheneau, I. Bloch, and S. Kuhr, Single-atom-resolved fluorescence imaging of an atomic Mott insulator, Nature (London) **467**, 68 (2010).

[53] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (John Wiley & Sons, Chichester, 2009), Vol. 405.

[54] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2009).

[55] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Integrating Neural Networks with a Quantum Simulator for State Reconstruction, Phys. Rev. Lett. **123**, 230504 (2019).

[56] E. Greplova, A. Valenti, G. Boschung, F. Schäfer, N. Lörch, and S. D. Huber, Unsupervised identification of topological phase transitions using predictive models, New J. Phys. **22**, 045003 (2020).

[57] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, and S. Kulik, Experimental neural network enhanced quantum tomography, npj Quantum Inf. **6**, 20 (2020).

[58] Y. Zhang, P. Ginsparg, and E.-A. Kim, Interpreting machine learning of topological quantum phase transitions, Phys. Rev. Research **2**, 023283 (2020).

[59] A. Bohrdt, S. Kim, A. Lukin, M. Rispoli, R. Schittko, M. Knap, M. Greiner, and J. Léonard, Analyzing Non-Equilibrium Quantum States through Snapshots with Artificial Neural Networks, Phys. Rev. Lett. **127**, 150504 (2021).

[60] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition* (Springer Science & Business Media, New York, 2013), Vol. 31.

[61] A. Valenti, G. Jin, J. Leonard, S. D. Huber, and E. Greplova, ManyBodyDynLearning, https://gitlab.com/QMAI/papers/manybodydynlearning (2021).

[62] I. López Gutiérrez and C. B. Mendl, Real time evolution with neural-network quantum states, Quantum **6**, 627 (2022).

[63] D. Luo, Z. Chen, J. Carrasquilla, and B. K. Clark, Probabilistic formulation of open quantum many-body systems dynamics with autoregressive models, arXiv:2009.05580.

[64] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, Tensorflow: A system for large-scale machine learning, in *Proceedings of the 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (2016), pp. 265–283.

[65] G. Casella and R. L. Berger, *Statistical Inference* (Duxbury, Pacific Grove, Belmont, 2002), Vol. 2.

[66] W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley & Sons, New York, 2008), Vol. 2.

[67] M. H. DeGroot and M. J. Schervish, *Probability and Statistics* (Pearson Education, Boston, 2012).

[68] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis* (CRC Press, Boca Raton, 2013).

[69] N. Choudhuri, S. Ghosal, and A. Roy, Bayesian methods for function estimation, Handb. Stat. **25**, 373 (2005).