Analyzing Nonequilibrium Quantum States through Snapshots with Artificial Neural Networks

A. Bohrdt,[1,2,3,4,*] S. Kim,[4] A. Lukin,[4] M. Rispoli,[4] R. Schittko,[4] M. Knap,[1,2] M. Greiner,[4] and J. Léonard[4]

[1]*Department of Physics and Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany*
[2]*Munich Center for Quantum Science and Technology (MCQST), Schellingstrasse 4, D-80799 München, Germany*
[3]*ITAMP, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA*
[4]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 23 January 2021; revised 11 August 2021; accepted 7 September 2021; published 7 October 2021)

Current quantum simulation experiments are starting to explore nonequilibrium many-body dynamics in previously inaccessible regimes in terms of system sizes and timescales. Therefore, the question emerges as to which observables are best suited to study the dynamics in such quantum many-body systems. Using machine learning techniques, we investigate the dynamics and, in particular, the thermalization behavior of an interacting quantum system that undergoes a nonequilibrium phase transition from an ergodic to a many-body localized phase. We employ supervised and unsupervised training methods to distinguish nonequilibrium from equilibrium data, using the network performance as a probe for the thermalization behavior of the system. We test our methods with experimental snapshots of ultracold atoms taken with a quantum gas microscope. Our results provide a path to analyze highly entangled large-scale quantum states for system sizes where numerical calculations of conventional observables become challenging.

DOI: 10.1103/PhysRevLett.127.150504

Introduction.—After a global quench in a thermalizing system, local observables approach a value that corresponds to their expectation value in a typical microcanonical many-body eigenstate of the system [1–3]. Depending on the properties of the system and the initial state, the path to thermal equilibrium can vary. For example, conserved quantities can slow down the equilibration process [4–6] or a quasistationary prethermal state can form, which exhibits properties different from the true thermal equilibrium state [7].

Quantum simulation experiments can enable the observation of the time evolution of a quantum many-body system starting from a nonequilibrium state with almost perfect isolation from the environment. In the past decade, a variety of nonequilibrium phenomena has been observed with examples ranging from exotic phases realized through Floquet driving [8–10] to many-body localization [11] and prethermalization [12]. In many cases, theory can provide a clear prediction as to which observables should be studied, such as a given order parameter for a well-known phase transition. For some problems, however, it is not as clear which observable to look at, and by making a choice for one specific quantity, valuable information might be discarded. In many platforms with microscopic readout, Fock space snapshots of the quantum many-body state are the measured dataset. Fock space snapshots provide a wealth of information about the quantum many-body state by providing access to both local observables and nonlocal, high-order correlations.

In order to address the challenge of finding suitable observables, artificial neural networks have recently emerged as a valuable tool in quantum many-body physics 17–13]] and in nonequilibrium statistical mechanics [18]. Previous machine learning approaches to study nonequilibrium systems have focused on quantities such as the entanglement spectrum [19–21] or full eigenstates [22], which are, however, experimentally inaccessible.

In this Letter, we study the dynamics of an interacting quantum many-body system in terms of experimental Fock space snapshots with the help of neural networks, Fig. 1(a). We find this analysis to have two main advantages: (i) These snapshots are directly measured in many quantum simulation platforms, and large numbers of snapshots can be routinely obtained. (ii) Raw data are used, where no analysis for specific quantities has taken place and all available information can be used without any bias. We consider the one-dimensional Bose-Hubbard model

$$\hat{\mathcal{H}} = \sum_i \left[-J(\hat{a}_i^\dagger \hat{a}_{i+1} + \text{H.c.}) + \frac{U}{2}\hat{n}_i(\hat{n}_i - 1) + W h_i \hat{n}_i \right]. \quad (1)$$

Here, $\hat{a}_i^{(\dagger)}$ annihilates (creates) a boson on site i and $\hat{n}_i = \hat{a}_i^\dagger \hat{a}_i$ is the particle number operator. The first term corresponds to hopping between neighboring sites, the second term is the interaction, here fixed at $U/J = 2.9$, and the last term is the quasiperiodic potential mimicking on-site disorder with amplitude W, which can be created in a cold atom setup with an incommensurate lattice

as $h_i = \cos(2\pi\beta i + \phi)$. In this Letter, we consider $1/\beta = 1.618$.

This system exhibits a many-body localized (MBL) phase, where thermalization breaks down as the disorder strength is increased beyond a critical value. The transition from an ergodic to a many-body localized phase is fundamentally different from the well-studied case of equilibrium phase transitions, as it describes a nonequilibrium setting [23–31]. Finding the transition point is numerically challenging, because it is usually obtained from entanglement properties or the level statistics, which can only be obtained for small system sizes where full diagonalization of the Hamiltonian is possible. Here, we focus on Fock space snapshots of the many-body quantum state as input data, which are the direct output of quantum gas microscopy experiments and thus experimentally readily accessible for the systems of interest. This approach has the advantage that significantly bigger system sizes can be reached experimentally.

We consider the dynamics of two one-dimensional systems of 8 and 12 sites, which are initialized in a Mott-insulating state with exactly one particle per site. In Fig. 1, we first train the network to distinguish snapshots of the many-body quantum state, obtained from exact diagonalization calculations, for low ($W/J = 0.3$) and high ($W/J = 11.0$) disorder strength for an interaction strength of $U/J = 2.9$ in the comparatively long-time limit at time $tJ = 100$. We average over ten different disorder realizations, obtained by varying the phase ϕ in the potential. After the network has learned to label the extremal cases correctly with sufficiently high accuracy ($> 90\%$), we input snapshots for intermediate values of the disorder strength. After training the neural network on numerically simulated snapshots, we use experimental data as input, where each snapshot stems from a different disorder realization. As output, for each disorder strength we obtain the fraction of snapshots labeled as "many-body localized" and "thermalizing," see Fig. 1. Based on these results, we conclude that the many-body localization transition is located within the range of $W/J \approx 4$–8 with strong finite-size drifts. This result is in agreement with previous experiments [32,33], which considered conventional observables such as the local entropy. Notably, the local entropy exhibits volume law scaling both in the thermal and the MBL phase and is thus by itself not sufficient to locate the transition without exact numerics [32]. Our results, in contrast, are able to distinguish the two phases without any theoretical input, which suggests that the network learned a more suitable observable to distinguish the two phases. In the Supplemental Material [34], we show the level statistics for system sizes $L = 6$–8 for comparison. Similar to the machine learning analysis of a disordered spin chain, based on the entanglement spectrum in [19], the transition found by the neural network is as sharp as the level statistics, but exhibits a small shift to larger disorder strengths.

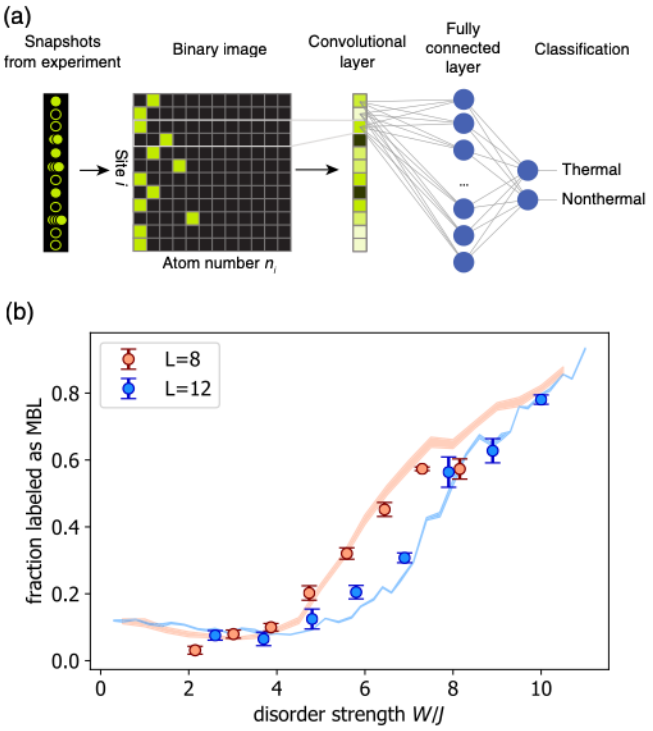

FIG. 1. Machine learning many-body localization. The Bose-Hubbard model with a quasiperiodic disorder potential exhibits a MBL phase, where thermalization breaks down, as the disorder strength is increased beyond a critical value. (a) We study the dynamics of the system after a quench for different disorder strengths by evaluating snapshots from a quantum gas microscope with neural networks. (b) A neural network is trained to distinguish exact diagonalization snapshots at $W/J = 0.3$ and $W/J = 11$ for $U/J = 2.9$ and a system with 8 and 12 sites at time $tJ = 100$ after a global quench. After the training process is finished, snapshots at intermediate values of the disorder strength are used as input. The plot shows the resulting classification for numerical data (shaded band) as well as experimental snapshots (symbols). As the system size is increased, the fraction of snapshots classified as MBL begins to increase at larger values of W, indicating the transition in the finite-size system. The accuracies are averaged over two independent runs and the errors denote one standard error of the mean (SEM).

While we have only compared two extremal disorder strengths in the long-time limit, the full dynamics of the system contain much more information. We proceed by analyzing the time- and disorder-strength dependence of the system after the global quench.

Learning thermalization.—We now investigate the system's approach to thermal equilibrium by comparing each time step to a thermal state of the same Hamiltonian. The performance of the network in distinguishing dynamics from equilibrium can then be used as a probe of thermalization.

In order to compare the time-evolved state to thermal equilibrium, all conserved quantities of the model should be considered [3]. In our experiment, both the energy

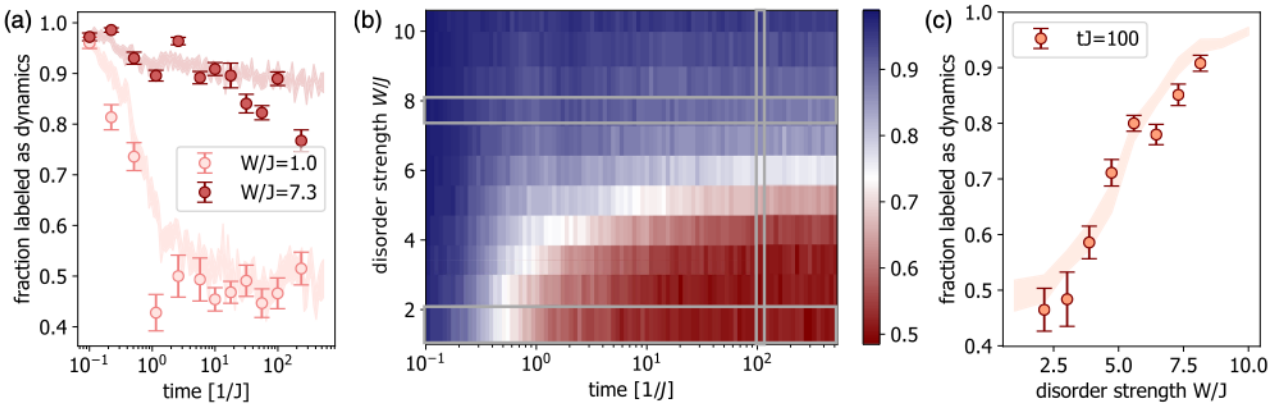

FIG. 2. Learning thermalization. A system with eight sites and $U/J = 2.9$ is initialized in a Mott-insulating state of one particle per site and the ensuing time evolution is investigated. In each time step, the neural network is trained to distinguish snapshots from the current time step from snapshots from a thermal state with the same energy density, both obtained from exact diagonalization. A high accuracy indicates that the current time step can be easily distinguished from the thermal state. (a) The resulting classification as "dynamics" versus "equilibrium" for $W/J = 1.0$ and $W/J = 7.3$, averaged over 12 different disorder realizations (shaded line). Experimental data from the dynamics after the quench is used as input at selected time steps (symbols). (b) Exact diagonalization results for disorder strengths between $W/J = 1$ and $W/J = 10$ for the full dynamics. (c) Classification as dynamics versus equilibrium at time $tJ = 100$ for disorder strengths between $W/J = 1$ and $W/J = 10$. The results are averaged over ten independent runs and the error bars correspond to the SEM.

density and the particle number are conserved during the many-body evolution. The energy density of the initial state is matched by choosing the temperature of the thermal state accordingly. We take the conservation of the total particle number into account by calculating the thermal state within a fixed particle number sector. We numerically generate snapshots from such a state in thermal equilibrium, as well as from the time-evolved state for each time step under consideration.

For each time step, we train the network to label the snapshots from the thermal equilibrium distribution as equilibrium and the snapshots from the numerically time-evolved initial state as dynamics. The neural network parameters are optimized for each time step separately. We then test the network's performance by inputting experimental data with different evolution times. In Fig. 2(a), the resulting classification into the categories dynamics versus equilibrium is shown as a function of time. Here, we average over 12 different disorder realizations and take snapshots at the corresponding effective temperatures.

For small W/J, the system thermalizes comparably fast: for times $tJ > 10$, the network reaches an accuracy of 50%, equivalent to guessing between the two classes. This means the network fails to distinguish snapshots from the time-evolved state from the corresponding thermal state. For high values of W/J, the system fails to thermalize on the timescales accessed here, and the network is able to distinguish the current time step from the thermal equilibrium state with a high accuracy. Using an interpretable network architecture [37], we find that, for intermediate disorder strengths, higher-order correlations play a role in the classification task (see the Supplemental Material [34]).

We study the long-time limit at $tJ = 100$ for a range of values of the disorder strength. As shown in Fig. 2(c), the fraction of snapshots classified as dynamics rises strongly between $W/J \approx 4$ and $W/J \approx 8$ and reaches values close to 1, indicating that the system has not reached thermal equilibrium.

We benchmark our experimental results by testing the network with theoretical snapshots not used during training and find good agreement throughout the range of the covered parameters.

This procedure has the advantage that the features used to make the classification can vary for different time steps and the network specifically searches for differences between the current time and thermal equilibrium. It is therefore, in principle, capable of identifying specific observables that have not yet reached their thermal equilibrium value and thus find, for example, (almost) conserved quantities. Indeed, with this method we find deviations from thermal equilibrium already in the range of $W/J \approx 2$–5, in contrast to the results from the classification scheme in Fig. 1(b). This indicates an improved sensitivity of our method. Here we consider a system that exhibits a transition from thermalizing behavior to many-body localization, which constitutes a canonical example in the study of nonequilibrium phenomena. Note, however, that our scheme is not limited to the system considered here and can be applied to a variety of models. This method also allows one to detect, for example, prethermal behavior and the existence of conserved quantities that keep their value during the dynamics and therefore never reach a generic thermal equilibrium value. Another canonical model to study equilibration behavior is the transverse field Ising

model, which has an extensive number of conserved quantities. In the Supplemental Material [34], we show that a neural network performs significantly worse in distinguishing the time-evolved state from an approximative generalized Gibbs ensemble, where a few conserved quantities are taken into account, than the simple thermal state discussed above, where only the energy density is considered. This highlights the capability of our approach to identify conserved quantities, which can drastically alter the thermalization process. Our method comes at the expense that one needs snapshots from the thermal density matrix for training, which—especially in the case of a nonthermalizing phase such as MBL—may need to be generated numerically. In the following, we overcome this limitation by analyzing the transition in the dynamics with an unsupervised scheme that, in principle, does not rely on theory data.

Confusion learning.—Several unsupervised learning schemes that use the network performance to probe whether and where a phase transition, or more generally, a qualitative change in the data, exists have been proposed [38–40]. Here, we adapt a scheme termed "confusion learning" introduced in Ref. [38]. In brief, the scheme works as follows: We have a dataset of snapshots for values of the disorder strength $0.3 \leq W/J \leq 11.0$. The goal is to test whether a value W^* exists at which the data change qualitatively. We start with a guess for W^* and label all snapshots for $W \leq W^*$ as phase A and correspondingly all snapshots with $W > W^*$ as phase B. Assuming the snapshots are qualitatively different for $W \leq W^*$ as compared to $W > W^*$, the network should achieve a high accuracy in assigning the correct labels. However, if there is no qualitative change at the W^* under consideration, there will be confusion about the correct labels and the accuracy will thus be lower. Therefore, if there is a qualitative change in the data, the accuracy as a function of W^* will be maximal if W^* corresponds to the transition point. Trivially, the test accuracy is expected to approach unity when the guessed W^* corresponds to the minimum or maximum value of W, because all data are labeled equally and no confusion occurs. In total, the presence of a critical point is therefore signaled by a characteristic W shape of the test accuracy as a function of the control parameter.

We train the neural network with numerical snapshots in the long-time limit ($tJ = 100$) in order to test for the presence of a phase transition. Subsequently, we use experimental data as input to the network [Fig. 3(a)]. The data show the onset of a maximum around $W^*/J = 7$, indicating the presence of a critical point in agreement with Fig. 1(b). The contrast in the W shape achieved here is comparable to the signal seen for a spin model in [38], where instead of snapshots the entanglement spectrum is used as input to the neural network. In order to isolate the signal of the phase transition from the trivial part of the W shape, we subtract the accuracy obtained when

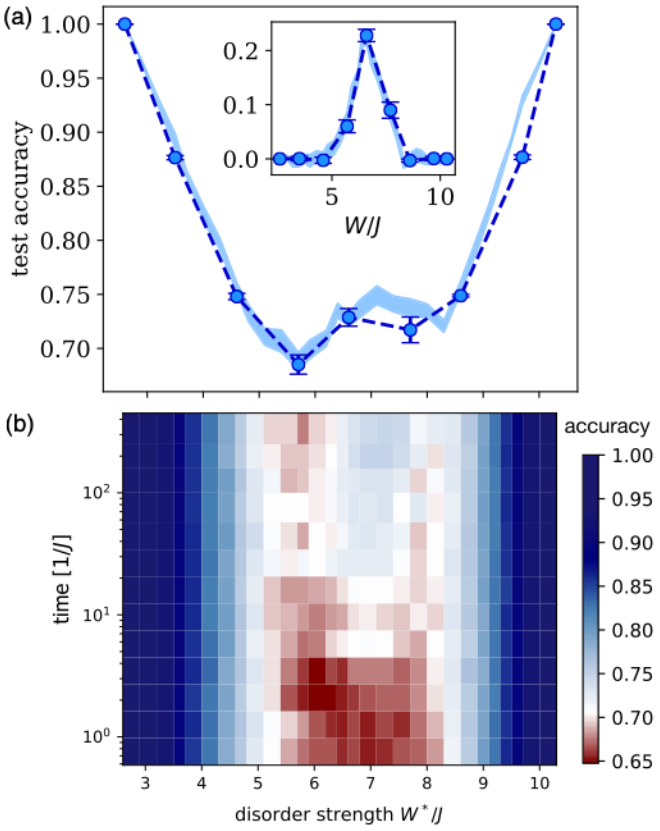

FIG. 3. Confusion learning. Snapshots of the many-body quantum state of a system with 12 sites, $U/J = 2.9$, and various disorder strengths W/J are analyzed using the confusion learning scheme. A neural network is trained to label all snapshots with $W < W^*$ as "phase A" and the remainder as "phase B." If a qualitative change in the data occurs, the accuracy will peak at an intermediate value of W^*. (a) The resulting accuracy at time $tJ = 100$ after the global quench for training on numerically simulated data (shaded line) and sorting experimental data (symbols). Inset: same data after subtracting the accuracy for randomly labeled data. (b) The accuracy for repeating the training process for different time points during the dynamics after the quench using numerically simulated data. The results are averaged over ten independent runs and the error bars correspond to the error based on one SEM.

training on randomly labeled data. The resulting difference, shown in the inset of Fig. 3(a), exhibits a clear peak at $W^*/J = 7$, that indicates the transition between the different dynamical phases. We also check with theoretical snapshots not used during training and find qualitatively similar behavior. We attribute the slight deviation in the maximum to the coarse resolution in the disorder strength for the experimental data.

Since differences in the thermalization behavior only present themselves in the course of the dynamics, we expect the phase transition to remain hidden at short evolution times. In order to reveal this effect, we perform the same method with theoretical snapshots at different evolution times. In Fig. 3(b), the resulting accuracy achieved by the network is shown as a function of W^*.

These results have several advantages compared to the previous methods: as opposed to Fig. 1(b), we do not *a priori* assume that there is a transition. Moreover, we specifically train the network to find differences between the snapshots at all available values of the disorder strength, thus avoiding bias from the choice of training data.

Summary and outlook.—In this Letter, we used machine learning techniques to study the nonequilibrium dynamics after a global quench in the one-dimensional Bose-Hubbard model with a quasiperiodic disorder potential. We used supervised as well as unsupervised machine learning methods to probe for a qualitative change in experimental snapshots as the disorder strength is tuned. Comparing the results for systems with 8 and 12 sites, we find that the critical value of the disorder strength increases with the system size, proving the need for methods applicable in large—experimentally accessible—systems. In contrast to standard tools to locate the MBL transition, the methods used here can be directly applied to experimental data taken with a quantum gas microscope and are not limited to small system sizes. We furthermore studied the approach to thermal equilibrium—or lack thereof—by training a neural network to distinguish snapshots from the current time step from snapshots from a thermal ensemble at the same energy and particle density. The accuracy achieved by the network indicates how nonthermal the time-dependent quantum many-body state is.

An exciting future research direction consists of applying the same scheme to identify conserved or almost conserved quantities in experimentally accessible data, for example, by using a generalized Gibbs ensemble for comparison. Apart from the concrete system studied here, it would be interesting to consider other models and phenomena, for example, quantum scars [41,42] and Hilbert space fragmentation [43–45]. In order to gain additional physical insights, interpretability is an extremely important direction for future work and it would be interesting to study which observables the network uses to make the classifications considered here [37] and how those observables change during the time evolution of the many-body system.

We would like to thank Eugene Demler, Fabian Grusdt, Florian Kotthoff, Cole Miles, and Frank Pollmann for fruitful discussions. We acknowledge support from the Technical University of Munich—Institute for Advanced Study, funded by the German Excellence Initiative and the European Union FP7 under Grant Agreement No. 291763, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC–2111–390814868, TRR80, DFG Grants No. KN1254/2-1 and No. KN1254/1-2, from the European Research Council (ERC) under the European Unions Horizon 2020 Research and Innovation Programme (Grant Agreement No. 851161), from the NSF Graduate Research Fellowship Program (S. K.), from the NSF Grants No. PHY-1734011, No. PHY-1806604l, and No. OAC-1934598, the Gordon and Betty Moore Foundations EPiQS Initiative, the Vannevar Bush Award, and the Swiss National Science Foundation (J. L.).

*Corresponding author.
bohrdt@fas.harvard.edu

[1] J. M. Deutsch, Quantum statistical mechanics in a closed system, Phys. Rev. A **43**, 2046 (1991).
[2] M. Srednicki, Chaos and quantum thermalization, Phys. Rev. E **50**, 888 (1994).
[3] M. Rigol, V. Dunjko, and M. Olshanii, Thermalization and its mechanism for generic isolated quantum systems, Nature (London) **452**, 854 (2008).
[4] J. Lux, J. Müller, A. Mitra, and A. Rosch, Hydrodynamic long-time tails after a quantum quench, Phys. Rev. A **89**, 053608 (2014).
[5] S. Mukerjee, V. Oganesyan, and D. Huse, Statistical theory of transport by strongly interacting lattice fermions, Phys. Rev. B **73**, 035113 (2006).
[6] A. Bohrdt, C. B. Mendl, M. Endres, and M. Knap, Scrambling and thermalization in a diffusive quantum many-body system, New J. Phys. **19**, 063001 (2017).
[7] J. Berges, S. Borsányi, and C. Wetterich, Prethermalization, Phys. Rev. Lett. **93**, 142002 (2004).
[8] 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).
[9] H. Miyake, G. A. Siviloglou, C. J. Kennedy, W. C. Burton, and W. Ketterle, Realizing the Harper Hamiltonian with Laser-Assisted Tunneling in Optical Lattices, Phys. Rev. Lett. **111**, 185302 (2013).
[10] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger, Experimental realization of the topological Haldane model with ultracold fermions, Nature (London) **515**, 237 (2014).
[11] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Luschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, Observation of many-body localization of interacting fermions in a quasirandom optical lattice, Science **349**, 842 (2015).
[12] M. Gring, M. Kuhnert, T. Langen, T. Kitagawa, B. Rauer, M. Schreitl, I. Mazets, D. A. Smith, E. Demler, and J. Schmiedmayer, Relaxation and prethermalization in an isolated quantum system, Science **337**, 1318 (2012).
[13] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).
[14] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin *et al.*, Netket: A machine learning toolkit for many-body quantum systems, SoftwareX **10**, 100311 (2019).
[15] B. S. Rem, N. Käming, M. Tarnowski, L. Asteria, N. Fläschner, C. Becker, K. Sengstock, and C. Weitenberg, Identifying quantum phase transitions using artificial neural networks on experimental data, Nat. Phys. **15**, 917 (2019).

[16] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. S. Davis, E. Khatami *et al.*, Machine learning in electronic-quantum-matter imaging experiments, Nature (London) **570**, 484 (2019).

[17] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, Classifying snapshots of the doped Hubbard model with machine learning, Nat. Phys. **15**, 921 (2019).

[18] A. Seif, M. Hafezi, and C. Jarzynski, Machine learning the thermodynamic arrow of time, Nat. Phys. **17**, 105 (2020).

[19] F. Schindler, N. Regnault, and T. Neupert, Probing many-body localization with neural networks, Phys. Rev. B **95**, 245134 (2017).

[20] J. Venderley, V. Khemani, and E.-A. Kim, Machine Learning Out-of-Equilibrium Phases of Matter, Phys. Rev. Lett. **120**, 257204 (2018).

[21] Y.-T. Hsu, X. Li, D.-L. Deng, and S. D. Sarma, Machine Learning Many-Body Localization: Search for the Elusive Nonergodic Metal, Phys. Rev. Lett. **121**, 245701 (2018).

[22] W. Zhang, L. Wang, and Z. Wang, Interpretable machine learning study of the many-body localization transition in disordered quantum Ising spin chains, Phys. Rev. B **99**, 054208 (2019).

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

[24] I. Gornyi, A. Mirlin, and D. Polyakov, Interacting Electrons in Disordered Wires: Anderson Localization and Low-T Transport, Phys. Rev. Lett. **95**, 206603 (2005).

[25] V. Oganesyan and D. A. Huse, Localization of interacting fermions at high temperature, Phys. Rev. B **75**, 155111 (2007).

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

[27] 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).

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

[29] M. Serbyn, M. Knap, S. Gopalakrishnan, and Z. Papić, N. Y. Yao, C. R. Laumann, D. A. Abanin, M. D. Lukin, and E. A. Demler, Interferometric Probes of Many-Body Localization, Phys. Rev. Lett. **113**, 147204 (2014).

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

[31] B. Chiaro *et al.*, Direct measurement of nonlocal interactions in the many-body localized phase, arXiv:1910.06024.

[32] 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).

[33] 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).

[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.150504 for more details, in particular, results on interpretability and the transverse field Ising model, taking into account conserved quantities of the model, including Refs. [35,36].

[35] M. Grady, Infinite set of conserved charges in the Ising model, Phys. Rev. D **25**, 1103 (1982).

[36] T. Prosen, A new class of completely integrable quantum spin chains, J. Phys. A **31**, L397 (1998).

[37] C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, Correlator convolutional neural networks: An interpretable architecture for image-like quantum matter data, Nat. Commun. **12**, 3905 (2021).

[38] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, Nat. Phys. **13**, 435 (2017).

[39] F. Schäfer and N. Lörch, Vector field divergence of predictive model output as indication of phase transitions, Phys. Rev. E **99**, 062107 (2019).

[40] 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).

[41] 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).

[42] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Weak ergodicity breaking from quantum many-body scars, Nat. Phys. **14**, 745 (2018).

[43] P. Sala, T. Rakovszky, R. Verresen, M. Knap, and F. Pollmann, Ergodicity Breaking Arising from Hilbert Space Fragmentation in Dipole-Conserving Hamiltonians, Phys. Rev. X **10**, 011047 (2020).

[44] V. Khemani, M. Hermele, and R. Nandkishore, Localization from Hilbert space shattering: From theory to physical realizations, Phys. Rev. B **101**, 174204 (2020).

[45] T. Kohlert, S. Scherg, P. Sala, F. Pollmann, B. H. Madhusudhana, I. Bloch, and M. Aidelsburger, Experimental realization of fragmented models in tilted Fermi-Hubbard chains, arXiv:2106.15586.