

Two-dimensional many-body localized systems coupled to a heat bath

Joey Li^{1,2}, Amos Chan,³ and Thorsten B. Wahl⁴

¹*Institute for Theoretical Physics, University of Innsbruck, 6020 Innsbruck, Austria*

²*Institute for Quantum Optics and Quantum Information of the Austrian Academy of Sciences, 6020 Innsbruck, Austria*

³*Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*

⁴*TCM Group, Cavendish Laboratory, Department of Physics, J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom*



(Received 24 October 2024; revised 2 June 2025; accepted 3 June 2025; published 30 June 2025)

We numerically investigate the effect of coupling a two-dimensional many-body localized system to a finite heat bath, using shallow quantum circuits as a variational ansatz. Specifically, we simulate optical lattice experiments with two components of ultracold bosons, where only one species is subject to a random disorder potential and the other acts as a heat bath. We obtain a filling fraction-dependent phase diagram with a critical filling consistent with experiments. We also perform long-time time-dependent variational principle (TDVP) simulations with matrix product states, but find that the quantum circuit approach produces superior results because of the lack of boundary effects. Finally, we present additional results on the two-point correlation functions and the quantum mutual information between sites.

DOI: [10.1103/PhysRevB.111.224211](https://doi.org/10.1103/PhysRevB.111.224211)

I. INTRODUCTION

The assumption that isolated systems act as their own heat baths and therefore reach thermal equilibrium was largely taken for granted until the beginning of this century, when several analytical [1,2] and numerical works [3,4] started to provide evidence for the existence of many-body localization (MBL). Experiments observing MBL [5–9] and a rigorous proof [10] in the last decade seemed to put the phenomenon on solid ground. However, in recent years, multiple works started questioning the existence of true MBL [11–13], even in one dimension. This would imply that MBL in one dimension is only a transient effect (as has long been conjectured for higher dimensions [14,15]), though on incredibly long timescales.

Extremely long thermalization times imply that strongly disordered isolated quantum systems behave many-body localized on experimentally relevant timescales. That is, for all practical purposes one can assume the existence of MBL, even in higher dimensions, where it has also been experimentally observed [16]. In this spirit, we investigate the effect of coupling a localized and delocalized system in higher dimensions, a subject that is relevant for the understanding of thermalization phenomena and their timescales. In particular, we study a two-dimensional “MBL-like” system coupled to a heat bath, and aim to reproduce the phase transition point found in two-dimensional optical lattice experiments with two-component ultracold bosons [17] as the size of the heat bath is increased. In those experiments, one boson species was subject to random disorder while the other only experienced the optical lattice potential, acting as a heat bath for the

former. The experiments were also theoretically analyzed in Ref. [18], albeit using a product state ansatz that assumes the observations were caused by glassy dynamics [19].

We use quantum circuits, a type of tensor networks, as a variational ansatz for our classical simulations. Shallow quantum circuits have proven useful for both the numerical description of one- [20–22] and two-dimensional MBL systems [23–25] and the rigorous classification of symmetry-protected topological MBL phases [26–28]. We optimize a two-dimensional shallow quantum circuit to approximately diagonalize a hardcore version of the Hamiltonian modeling the experiments of Ref. [17]. The unitaries building up the quantum circuit contain the variational parameters of the approach. After optimizing those, we have an approximate representation of the overall unitary diagonalizing the Hamiltonian, and can extract local features (such as correlations) of its eigenstates. We analyze entanglement features in the different particle sectors of the diagonalizing unitary to obtain the phase diagram as a function of the disorder strength and the fraction of bosons not experiencing the disorder potential. [Note that we use the semantics of a phase transition throughout this work, although the transience of MBL implies that instead we have a (relatively sharp) crossover between a regime of extremely slow, experimentally inaccessible thermalization and fast thermalization.] This phase diagram constitutes the central novel finding of our work. We obtain a phase transition point in the same regime as the measured value.

For comparison, we also simulated real-time evolution with matrix product states using the time-dependent variational principle [29,30]. We found that the requirement of numerically accurate time evolutions restricts us to fairly small system sizes (6×6) with open boundary conditions. As a result, boundary effects and the shift of the two-dimensional MBL-to-thermal transition point with system size [15,30] lead to significantly lower critical disorder strengths. However, the

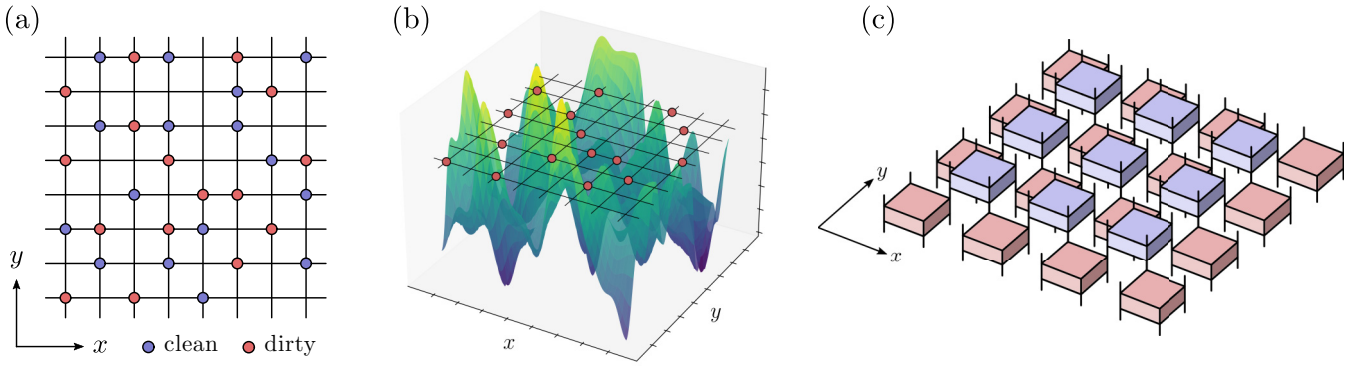


FIG. 1. (a) Optical lattice filled with two species of bosons: “clean” (blue) and “dirty” (red). Half the lattice sites are occupied: the overall filling is $\nu = 0.5$. We consider various relative fractions f_c and f_d of clean and dirty particles: depicted here is the case $f_c = f_d = 0.5$. (b) The system is in a random potential that affects *only* the dirty particles. (c) Two-layer quantum circuit U used as a variational ansatz to approximately diagonalize the Hamiltonian. (In this work we consider system sizes $L = 6$ and $L = 10$.)

corresponding phase diagram is qualitatively the same as the one obtained with quantum circuits, which do not suffer from these restrictions (but rather finite-entanglement effects). We thus find that our method is more suitable to describe the limit of large, experimentally relevant system sizes.

Furthermore, we extracted the correlation lengths from our optimized unitaries, which show a slow decay with disorder strength. Finally, we find that the quantum mutual information distributions between sites separated by short and medium distances develop longer tails as the disorder strength is lowered towards the transition point. If similar features are found in future, more accurate studies, then resonances are also at play on experimentally relevant timescales near the transition in higher dimensions, similar to one-dimensional MBL [31].

II. MODEL AND NUMERICAL APPROACH

We study a two-dimensional MBL system coupled to an immersed heat bath. Specifically, we consider two components of bosons confined by an optical square lattice. One species of bosons (“dirty bosons”) experiences a randomly chosen on-site disorder potential while the other one (“clean bosons”) acts as a heat bath. We assume that the dynamics of the bosons is governed by the Hamiltonian of Ref. [17],

$$H_{\text{exp}} = -J \sum_{\langle i,j \rangle, \sigma} a_{i,\sigma}^\dagger a_{j,\sigma} + \frac{V}{2} \sum_{i,\sigma} n_{i,\sigma} (n_{i,\sigma} - 1) + V \sum_i n_{i,d} n_{i,c} + \sum_i \delta_i n_{i,d}, \quad (1)$$

where $i = (x, y)$ denotes the lattice site. J is the tunneling amplitude, i.e., hopping strength, V the on-site repulsion strength, and δ_i the random disorder potential, which is chosen according to a Gaussian distribution with half-maximum width Δ . We consider only bosons in the center of the optical trap, where the trap potential can be neglected. $a_{i,\sigma}$ denotes the annihilation operator of a boson at site i for $\sigma \in \{c, d\}$ with c (d) corresponding to clean (dirty) bosons and $n_{i,\sigma} \equiv a_{i,\sigma}^\dagger a_{i,\sigma}$ the corresponding particle number operator. See Fig. 1 for an illustration. We define f_c as the relative fraction of clean bosons, and f_d as that of dirty bosons, where $f_c + f_d = 1$.

In numerical simulations, one typically has to truncate the on-site occupation number of bosons, which can be, in principle, arbitrarily large. As experimentally only doublon fractions of up to 18% are observed [17] (owing to the large on-site repulsion of $V = 24.4J$) and to make our simulations numerically feasible, we restrict ourselves to hardcore bosons. Our Hamiltonian becomes

$$H = -J \sum_{\langle i,j \rangle, \sigma} a_{i,\sigma}^\dagger a_{j,\sigma} + \sum_i \delta_i n_{i,d}. \quad (2)$$

We note that because of the (infinite) on-site repulsion, this Hamiltonian still describes an interacting model. We work with an $L \times L = 10 \times 10$ system with periodic boundary conditions. (We also consider, for comparison, the case $L = 6$.)

A. Numerical methods

Our study involves variationally diagonalizing the Hamiltonian (2), i.e., obtaining an approximation of the full set of eigenstates. (Independently, we also simulate time evolution using TDVP; this forms the basis for the results in Sec. III B.)

We use an ansatz U that is a two-layer tensor network, or quantum circuit [23–25], with the structure as depicted in Fig. 1(c). The ansatz is essentially a two-dimensional matrix product operator (MPO), or a generalization of projected entangled pair states (PEPS) that contains all eigenstates.

We numerically optimize U such that $UHU^\dagger$ is approximately diagonal. To that end, we minimize a cost function that is a sum of local tensor network contractions. Since the Hamiltonian (2) preserves the particle numbers of both clean and dirty bosons, its eigenstates also have fixed clean and dirty particle numbers. We impose this at the level of the individual unitaries building up the quantum circuit in our numerical approach [23,24]. This makes it straightforward to access individual particle sectors of our overall unitary. For further details about the ansatz, see Appendix A.

Cost function

We perform the variational diagonalization by minimizing a cost function, which can be motivated by the concept of local integrals of motion (LIOMs) [32–40], also known as 1-bits. MBL systems are known to possess a set of LIOMs $\{\tau_i\}$, which

(in two dimensions approximately [41]) commute with the Hamiltonian and each other, i.e., $[H, \tau_i] = [\tau_i, \tau_j] = 0$. Each τ_i is an operator localized around site i , and is defined by $\tau_i = U_{\text{exact}} \sigma_i^z U_{\text{exact}}^\dagger$ for the case of spin-1/2 spin models such as the “standard” disordered Heisenberg model [3,4,42], or analogously otherwise. U_{exact} is the local unitary operator that diagonalizes the Hamiltonian. Here, we define a set of *approximate* LIOMs given by $\tilde{\tau}_{i,\mu}(U) = U n_{i,\mu} U^\dagger$, where $\mu = c, d$, and U is the variational quantum circuit. To optimize U , we define a cost function $f \propto \sum_{i,\mu} \|[H, \tilde{\tau}_{i,\mu}]\|^2$ that penalizes the noncommutation of H and $\tilde{\tau}_{i,\mu}$. (The $\tilde{\tau}_{i,\mu}$ already commute with each other by construction.) Thus, by minimizing the cost function we are optimizing the approximate LIOMs. We note that although the motivation was physical, it is also possible to mathematically justify this choice of cost function by showing from the definition of $\tilde{\tau}_{i,\mu}$ that $[H, \tilde{\tau}_{i,\mu}] = 0$ for all $\tilde{\tau}_{i,\mu}$ if and only if $U H U^\dagger$ is diagonal.

Specifically, the cost function $f(U)$ is given by

$$\begin{aligned} f(U) &= \frac{1}{2L^2 3^{L^2}} \sum_i \sum_{\mu=c,d} \text{Tr}([H, \tilde{\tau}_{i,\mu}]^\dagger [H, \tilde{\tau}_{i,\mu}]) \\ &= \frac{1}{L^2 3^{L^2}} \sum_i \sum_{\mu=c,d} [\text{Tr}(H^2 (\tilde{\tau}_{\mu,i})^2) - \text{Tr}((H \tilde{\tau}_{i,\mu})^2)], \end{aligned} \quad (3)$$

where the factor of $1/2L^2 3^{L^2}$ has been included so that $f(U)$ is an $\mathcal{O}(1)$ number. Importantly, $f(U)$ is a sum of quantities that only involve local causal cone calculations, so the total computational cost scales linearly with the system size. We illustrate the tensor network contraction needed to calculate the cost function in Fig. 7(a) in Appendix A. For further details on the calculation, we refer the reader to Refs. [23,24], where a similar optimization procedure was employed.

B. Approximate eigenstates

We note that the approximate eigenstates encoded in our unitary U represent the dynamics of the Hamiltonian H on short timescales: it does not make a significant difference whether we consider the exact evolution of an initial state $|\psi_0\rangle$ via $|\psi(t)\rangle = e^{iHt} |\psi_0\rangle = U_{\text{exact}} e^{iE_{\text{exact}} t} U_{\text{exact}}^\dagger |\psi_0\rangle$ or the approximate evolution $|\tilde{\psi}(t)\rangle = U e^{iEt} U^\dagger |\psi_0\rangle$, as long as the time t is small, that is, less than $1/\|[H, \tilde{\tau}_i^z]\|_{\text{op}}$. (Here, $E_{\text{exact}} = U_{\text{exact}}^\dagger H U_{\text{exact}}$ is the diagonal matrix of the exact energies, and E is the diagonal matrix of the variational energies of the approximate eigenstates encapsulated in U .) As our approximate local integrals of motion $\tilde{\tau}_i^z$ do not commute with the Hamiltonian exactly, they do not persist to infinitely long times, but decohere on a timescale of the order of $1/\|[H, \tilde{\tau}_i^z]\|_{\text{op}}$ [41]. We note that, given the relatively small causal cones of our ansatz, the timescales we capture are shorter than the experimental ones. Therefore, absent other errors, the transition points we obtain will be lower bounds on the experimental ones (as for any disorder strength, localization might become unstable if we could consider longer timescales). Nonetheless, in our previous work using quantum circuits, we found excellent quantitative agreement between the experimental and theoretical transition point [24]. We also note that, although atom loss

is present in the experiment [17], it is not relevant on the short timescales where our approximation is valid.

Once the optimized unitaries have been obtained, we can calculate one- and two-point functions of the approximate eigenstates (see Appendix B). For the purpose of producing the phase diagram, we estimate the location of the transition (technically, crossover) point by using the fact that the variation of entanglement between different disorder realizations is greatest at the critical disorder strength [43]. Specifically, the process is as follows: One can extract the on-site entanglement entropies with respect to sites i . After averaging over eigenstates in particle sectors and calculating the fluctuation with respect to disorder realizations, one obtains maxima around the phase transition point [21,22,24,25,43]. (To improve numerical stability, in the final step, one also averages the entanglement entropy fluctuation over sites i .) The disorder strengths Δ where these maxima are located can be used to extract the phase transition points. We carry out this analysis for an overall filling fraction $\nu = 0.5$ as in the experiment [17] and as a function of the fraction $f_c \in [0, 1]$ of clean bosons, by explicitly selecting approximate eigenstates with the desired values of ν and f_c . For our simulations, we consider the Hamiltonian as defined on a 10×10 lattice with periodic boundary conditions, i.e., we have $N_c + N_d = 50$ bosons.

III. RESULTS

Here, we present our findings on the phase diagram as a function of the fraction f_c of clean particles along with a comparison to analogous results from simulation of real-time evolution. We then show results on correlation functions and quantum mutual information.

A. Clean particle fraction-resolved phase diagram

In Fig. 2, we show the numerical data obtained by optimizing our quantum circuit ansatz for 30 disorder realizations as a function of disorder strength Δ . To increase numerical stability, we used the same 30 disorder profiles for all values of Δ and only varied the overall prefactor [21,23,24]. As detailed above, we calculated the fluctuation of the eigenstate-averaged (sampling over 4000 approximate eigenstates) on-site entanglement entropy with respect to disorder realizations and averaged the result over the site positions. We restricted the eigenstates to have $N_c + N_d \equiv N_{\text{tot}} = 50$ bosons with $N_c/N_{\text{tot}} = f_c$ fixed. The corresponding filling fraction-resolved entanglement entropy fluctuation is shown in Fig. 2. (We note that we did perform an error analysis, using the method described in Ref. [24]—the resulting error bars are smaller than the scatter points and therefore not visible.) For all fractions f_c of clean particles, we find a unique maximum in the entanglement entropy fluctuation (see inset of Fig. 2), whose disorder strength value we use to plot the filling-fraction dependent MBL-thermal phase diagram. We note that for disorder strengths $\Delta \gtrsim 30J$, our phase boundary deviates from the expected behavior of approaching a horizontal line asymptotically. Such a behavior is required, as the eigenstates for any arbitrarily large Δ are always delocalized if f_c is close enough to 1, such that the $N_d = (1 - f_c)N_{\text{tot}}$ dirty bosons cannot localize the clean ones. Finding the transition point for

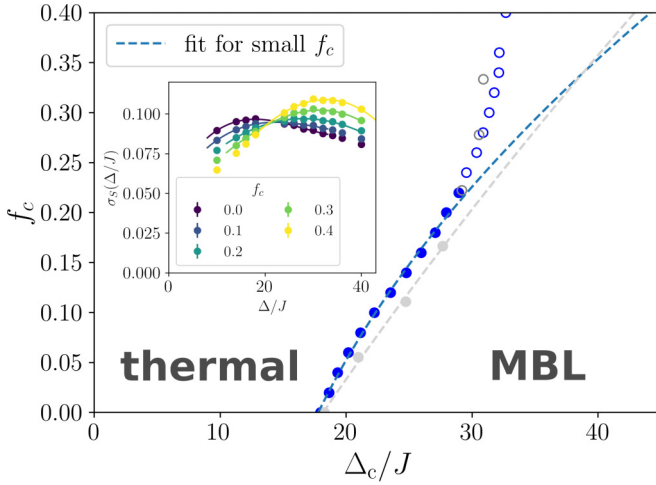


FIG. 2. Phase diagram of thermal and MBL-like regimes for overall filling fraction $\nu = 0.5$ obtained by extracting the transition points from the maxima of the entanglement entropy fluctuation for different f_c in the inset. The blue-dashed line indicates a second order polynomial fit to our data for small f_c (blue-filled circles), from which our method deviates for $\Delta \gtrsim 30J$ (blue empty circles). The gray symbols indicate the corresponding results for a 6×6 system, see Appendix for details. (Inset) Entanglement entropy fluctuation (see main text) as a function of disorder strength Δ with fractions f_c of clean particles as indicated in the legend. The error bars are smaller than the symbol size.

large disorder strengths becomes increasingly difficult for our method, as there are fewer and fewer dirty particles that must localize the clean particles over larger and larger length scales in order to remain MBL. Our short-depth quantum circuit ansatz is unable to capture such large localization lengths and a breakdown of our approach for fractions of clean particles close to 1 is therefore expected.

Because of the asymptotic flattening of the phase boundary, the Taylor expansion of $f_c(\Delta_c)$ should have a negative coefficient in second order. We fit our data to such a second order polynomial in the regime where the data points follow a curve that is right-bent. Crucially, the transition point $\Delta_c(f_c = 0) = 18J$ is consistent with Ref. [23], i.e., the small f_c results are reliable. Specifically, they match the expected (fitted) behavior in the range $0 \leq \Delta_c(f_c) \leq 30J$, which covers the value of $\Delta = 28J$ used in the experiment [17]. For the experimental disorder strength, we find a transition at $f_c(\Delta = 28J) \approx 0.20$, which compares favorably with the experimental value of $f_c^{\text{exp}} = N_c/N_{\text{tot}} \approx (40 \pm 5)/(124 \pm 12) \approx 0.33 \pm 0.07$ [17] given the simplicity of our approximation. The still significant deviation from the experimental value is mostly caused by the limited entanglement in the ansatz, as well as the truncation of the local Hilbert space. Specifically, the unitaries building up our quantum circuit act only on plaquettes of 2×2 sites, which limits the amount of entanglement allowed by our ansatz, and we considered a hardcore Bose-Hubbard model, i.e., neglected double and higher occupancies. Mitigating either of these restrictions would likely lead to a transition point closer to the experimental one, but would demand vastly larger computational resources. (Our theoretical result of $f_c(\Delta = 28J) \approx 0.20$ has negligible error

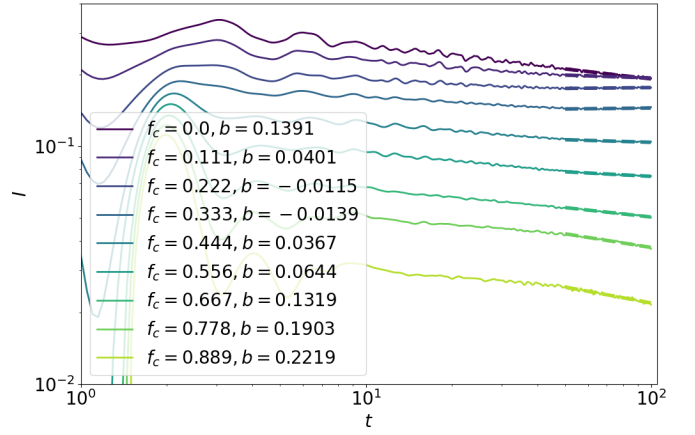


FIG. 3. Imbalances I as a function of time t for different fractions $f_c = N_c/18$ of clean particles obtained for a 6×6 system initialized with a charge density wave of $N_c + N_d = 18$ bosons. The data are averaged over ten disorder realizations. The dashed lines indicate a polynomial fit $I = at^{-b}$ for late times ($a, b \in \mathbb{R}$). Starting from large f_c , we find that the exponent b vanishes at around $f_c \approx 0.4$. For small f_c , b is positive again, as the initial exponential decay of the imbalance has not fully subsided yet.

bar, but the error analysis only takes into account statistical error caused by disorder realizations and eigenstate sampling, not the systematic errors just mentioned.)

We also note that the transition point f_c , from numerical simulations of a 10×10 system, is closer to the experimental value than the one found for a 6×6 system, indicating the correct trend (cf. Appendix C). This underpins the statement that, owing to their local structure, shallow quantum circuits are capable of capturing the transition points found with charge-density-wave imbalance experiments [24].

B. Real-time dynamics

Complementing our variational diagonalization, we also simulated time evolution using the time-dependent variational principle (TDVP) on matrix product states (MPS) [29], following Ref. [30]. TDVP simulations were done using the TeNPy library [44]. Specifically, we consider a 6×6 system with open boundary conditions and initialize it in a charge-density-wave state with alternating full and empty columns. The full columns are occupied with one boson per site—randomly either a dirty or a clean atom—such that the overall number of clean particles is N_c and the corresponding fraction of clean particles is $f_c = N_c/18$. We performed the evolution under the hardcore Hamiltonian Eq. (2) and calculated the imbalance defined as $I(t) = \frac{N_e(t) - N_o(t)}{N_e(t) + N_o(t)}$ at each time step, where $N_{e/o}(t)$ is the total number of atoms in all even/odd columns. We found that for a maximum MPS bond dimension of $\chi = 384$, and a time step size of $\Delta t = 0.05$, the results for the imbalance are sufficiently well converged, even for long evolution times of $t = 100J^{-1}$ hoppings.

We present the results for the imbalance as a function of time and for the example of disorder strength $\Delta = 8J$ and different fractions f_c of clean particles in Fig. 3. The dashed lines in the plots indicate polynomial fits whose exponents can be used to determine the critical value of f_c for a given

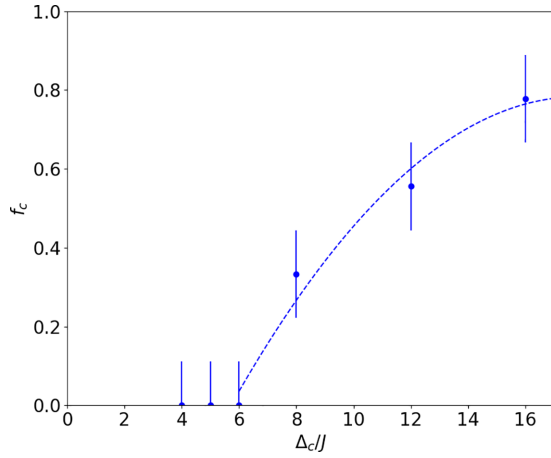


FIG. 4. Phase diagram as a function of the fraction f_c of clean particles and disorder strength Δ obtained from the imbalance evolutions for different values of Δ (cf. Fig. 3 for $\Delta = 8J$). We averaged over ten disorder realizations for each disorder strength. Error bars denote the uncertainty in determining at which fraction f_c the exponent b vanishes. The dashed line is a guide to the eye.

disorder strength (as the fraction f_c where the exponent of becomes numerically zero) [30]. Afterwards, we used the critical fractions f_c for different disorder strengths to infer the filling fraction-dependent phase diagram, shown in Fig. 4. Compared with Fig. 2, we notice that the critical disorder strengths for fixed f_c are shifted to significantly lower values, which are incompatible with the experimental transition point of $f_c^{\text{exp}} \approx 0.33 \pm 0.07$. We ascribe this to strong boundary effects (20 sites of 36 are next to the edge) and the two-dimensional MBL-to-thermal transition points being at significantly lower disorder strengths for smaller systems [15,30]. However, we note that the phase diagram obtained with TDVP is qualitatively similar to the one obtained with quantum circuits. In the former case, we are unfortunately unable to reach much larger system sizes because of limitations of computational resources. (The bond dimensions required for sufficient accuracy quickly get too large.) Furthermore, closing the boundaries would lead to additional, spurious delocalization not present in the experimental system. We thus conclude that quantum circuit simulations are better suited than TDVP to reproduce the two-dimensional MBL-to-thermal transition points of the considered two-species Bose-Hubbard Hamiltonian, at least for experimentally relevant system sizes. Nonetheless, the TDVP simulations qualitatively support our quantum circuit results.

C. Two-point correlation function

We now return to the discussion of results obtained from the quantum circuit study. Our method allows us to use the optimized unitaries to calculate two-site reduced density matrices (see Appendix B for details), from which we can compute two-point correlation functions and the quantum mutual information (QMI) between sites.

We first consider correlation functions. Within our approximation, the two-site correlation function $C_{\sigma\mu}(i, j) = \langle n_{i,\sigma} n_{j,\mu} \rangle - \langle n_{i,\sigma} \rangle \langle n_{j,\mu} \rangle$, with $\sigma, \mu \in \{c, d\}$, is nontrivial only

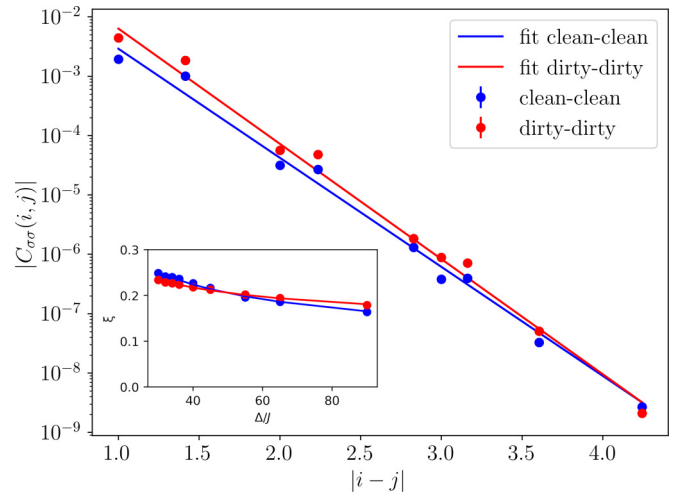


FIG. 5. Decay of correlations as a function of distance $|i - j|$, shown for the example of $\Delta = 36J$. The data were obtained by averaging over 100 randomly chosen approximate eigenstates with fraction $f_c = 0.2$ of clean bosons and averaging over 30 disorder realizations. Error bars denote fluctuations between disorder realizations and are smaller than the symbol size. For the fits, we used linear functions after taking the logarithm for the three data points of smallest distance. (Inset) Correlation lengths extracted from the fits as shown in the main figure for various values of Δ and $f_c = 0.2$. The solid lines denote $1/\ln(\Delta/J)$ fits [45–47].

in a fixed range of separations between sites i and j (the causal cone). The obtained correlations and extracted correlation lengths are shown in Fig. 5. We note the slow decay of the average correlation lengths with disorder strength Δ , cf. Ref. [23]. Furthermore, we find that the correlation lengths for dirty and clean particles are very similar, indicating that the dirty particles are effective at localizing the clean ones in the MBL regime within our approximation. For example, for $f_c = 0.2$ we observe that $\xi_{cc}(\Delta = 45J) = 0.22 \pm 0.01$ and $\xi_{dd}(\Delta = 45J) = 0.21 \pm 0.01$, decaying to $\xi_{cc}(\Delta = 90J) = 0.16 \pm 0.01$ and $\xi_{dd}(\Delta = 90J) = 0.18 \pm 0.01$. Our predictions for the two-point correlations can be compared to future optical lattice experiments.

D. Quantum mutual information

Finally, we consider the quantum mutual information (QMI), which has been used to quantify correlations in MBL systems [48,49], and more recently has been studied in the context of resonances [50]. Here, we analyze the distribution of the QMI between individual sites in our two-dimensional model (2). The QMI of an eigenstate $|\psi\rangle$ is defined as

$$I(A : B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}), \quad (4)$$

where the system is partitioned into three disjoint parts A , B , C , and the QMI is between regions A and B . ρ_X is the reduced density matrix of $|\psi\rangle$ obtained by tracing out its complement $\bar{X}$,

$$\rho_X = \text{tr}_{\bar{X}}(|\psi\rangle\langle\psi|), \quad (5)$$

and $S(\rho_X) = -\text{tr}(\rho_X \ln(\rho_X))$ denotes the von Neumann entropy.

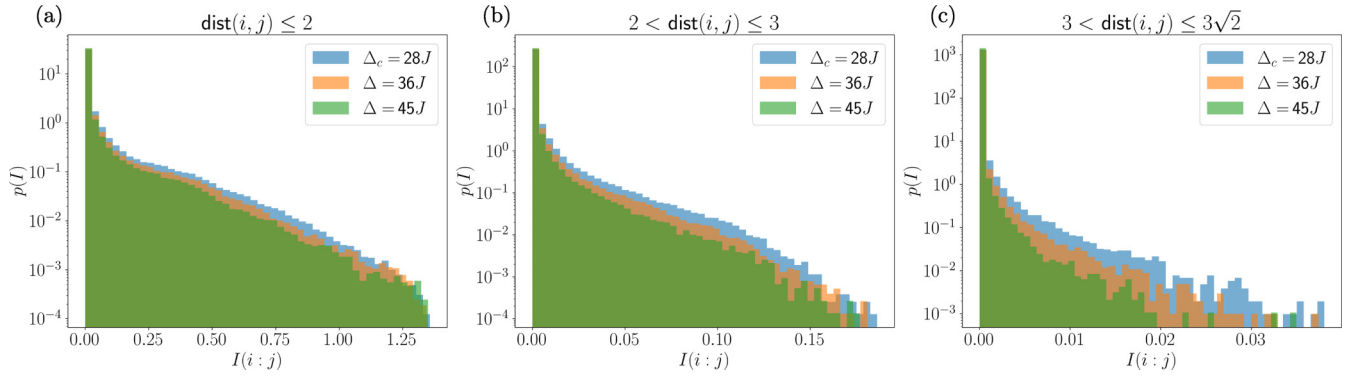


FIG. 6. Probability density of the two-site QMI obtained by sampling over 100 eigenstates and 30 disorder realizations, taking all sites into account, for $f_c = 0.2$ (corresponding to a transition point $\Delta_c = 28J$) and the values of Δ indicated in the legend. The distances between the sites are $\text{dist}(i, j) \leq 2$ in (a), $2 < \text{dist}(i, j) \leq 3$ in (b), and $3 < \text{dist}(i, j) \leq 3\sqrt{2}$ in (c).

We first review the expected behavior of the eigenstate QMI probability distribution. In general, the two-site QMI $I(i : j)$ decays exponentially with distance in the MBL regime, and decays more slowly in the thermal regime [48]. The probability distribution of values of the two-site QMI in thermal-regime eigenstates is roughly symmetric and centered around a value independent of the distance $|i - j|$, while in the MBL regime, the QMI distribution is more one-sided with greatest probability near zero, but also longer tailed, with more instances of near-maximal QMI, particularly for small $|i - j|$ [49]. When considering the QMI distributions at various disorder strengths, the distribution is expected to be independent of disorder strength for neighboring sites, while for large $|i - j|$ the distribution is more heavily centered around 0 at very strong disorder, and longer tailed closer to the transition point.

Using our variational method, we calculate the QMI from the two-site reduced density matrices (which can be obtained as illustrated in Appendix B), sampling over 100 approximate eigenstates per disorder realization and disorder strength. As our ansatz only captures area law-entangled states, we are unable to study the QMI distribution in the thermal regime. The results, for disorder strengths at the transition point and above, are shown in Fig. 6. We note that, although the local Hilbert space dimension of this model is 3, the maximum QMI between two sites is $I(i : j) = 2 \ln 2$, equivalent to that of two maximally entangled qubits, as particle number conservation prohibits the existence of maximally entangled two-qutrit states on a subsystem of two sites.

As our approximation only captures nontrivial correlations within a fixed range of distances, we can only consider certain separations between sites i and j . We consider the short-ranged, i.e., nearest-neighbor case in Fig. 6(a), where the distributions are roughly similar for different disorder strengths and show instances of near-maximal QMI, as expected based on analogous studies in one-dimensional MBL. We also consider the “medium-ranged” case in Figs. 6(b) and 6(c). We are unable to consider the long-ranged case, as within our approximation, sites are uncorrelated for $|i - j| > 3\sqrt{2}$. In addition, a further limitation arising from the local nature of the ansatz is that the QMI is consistently $\lesssim 0.2$ for $|i - j| \geq 2$ and $\lesssim 0.05$ for $|i - j| > 3$, independently of the

model under consideration. Thus in Figs. 6(b) and 6(c), the observed values of the QMI are likely significantly below the true values. However, the qualitative aspects of these plots may still capture some of the relevant physics. This is because each approximate eigenstate is essentially the projection of an exact eigenstate onto a manifold of weakly entangled states; since the projection is the same in all cases, it is still meaningful to consider *relative* differences, especially correlations or related quantities between different sets of eigenstates. In particular, this allows us to compare the two-site QMI, at fixed distances, between systems at different disorder strengths.

We now discuss the relation to resonances, which have been suggested to be the driving force behind the transition from the MBL to the thermal regime in one dimension [31, 50–53]. Resonances can be viewed as superpositions of few (deep in the MBL regime, two) near-product states whose local occupations differ in many sites, e.g., cat states. A crucial requirement is that those superpositions are the result of the near-product states having hybridized owing to them being close in energy. This can be detected via the QMI if the regions A and B (in our case, sites i and j) are far away (typically of the order of the system size), such that a nonvanishing QMI indicates the superposition of macroscopically distinct near-product states, which, because of the locality of the Hamiltonian, must be close in energy [50]. Because of the locality of our ansatz, we are unable to detect such long-range resonances. Instead, we study how the QMI between the most distant sites accessible builds up as the disorder strength Δ is lowered towards the transition.

We observe that in Fig. 6(c) the distribution of the QMI develops tails near the transition that are absent for larger Δ . We note that the lengths of these tails are constrained owing to the aforementioned limitations arising from the short-range nature of the ansatz. We do not observe the near-maximal QMI that is the hallmark of many-body resonances, in the “medium-ranged” case. However, the qualitative behavior of the distribution is reminiscent of the expected behavior of the distribution of the longer-range QMI as the disorder strength is lowered towards the transition point. While we do not capture resonances with our numerical methods, within the range of QMI and site separations that we do capture, we observe a broadening in the QMI distribution with decreasing Δ that

corresponds to the increased presence of resonances near the transition point, analogous to the one-dimensional case.

IV. CONCLUSIONS

We used quantum circuits to simulate a two-component ultracold bosonic gas, where one component is localized by a disorder potential and the other acts as a heat bath. We found a transition point of the fraction of clean particles in the same regime as the experiments of Ref. [17]. We also calculated the correlation lengths as quantities to be compared to future experiments. Finally, we computed the QMI of our approximate eigenstates; we found that the distribution of the QMI gets broader as the disorder strength is decreased towards the transition. These results support the perspective that the proliferation of resonances contributes to the two-dimensional MBL-to-thermal crossover seen on experimental timescales—analogue to the case in one dimension. Further studies are required to confirm this hypothesis.

ACKNOWLEDGMENTS

We thank David Huse and Hannes Pichler for helpful discussions. J.L. acknowledges support from ERC Starting grant QARA (Grant No. 101041435) and the Austrian Science Fund (FWF): COE 1 and quantA. A.C. acknowledges support from the Royal Society Grant No. RGS\R1\231444 and from the EPSRC Open Fellowship EP/X042812/1. T.B.W. acknowledges funding from an EPSRC ERC underwrite Grant No. EP/X025829/1. The computational results presented here have been achieved (in part) using the LEO HPC infrastructure of the University of Innsbruck.

APPENDIX A: VARIATIONAL ANSATZ

In this work, we use the quantum circuit U , a two-layer tensor network with the structure depicted in Fig. 1(c), to variationally diagonalize the Hamiltonian (2), i.e., approximately calculate the full set of eigenstates. [Only a subset of the tensor network is depicted in Fig. 1(c) for clarity; since we

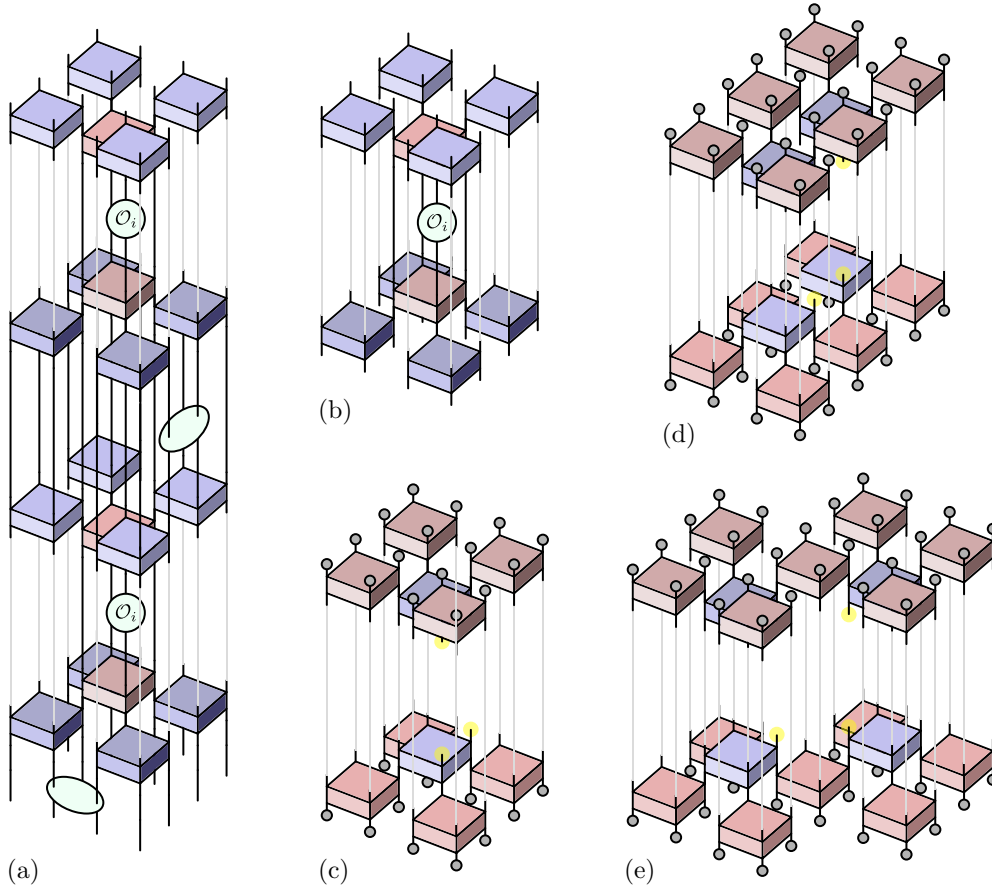


FIG. 7. (a) Illustration of the summands that appear in $\text{Tr}((H\tilde{\tau}_{i,\mu})^2)$ in the cost function given in Eq. (3), where the ovals illustrate nearest-neighbor terms in H . Note that the trace would correspond to the open legs at the top and the bottom being identified. (b) Illustration of $U^\dagger \mathcal{O}_i U$ of a local operator $\mathcal{O}_i$ acting on site i . Using unitarity, $U^\dagger \mathcal{O}_i U$ has nontrivial support only in the causal cone of operator $\mathcal{O}_i$. (c)–(e) Diagrammatic representations of the two-site reduced density matrix $\rho_{ij} = \text{Tr}_{\overline{ij}}(|\psi\rangle\langle\psi|)$. Different choices of the sites i and j lead to causal cones of size (c) 4×4 , (d) 4×6 , and (e) 28 sites. Note that these diagrams are inverted: the top tensor network is $\langle\psi|$ and the bottom tensor network is $|\psi\rangle$. For clarity, the tensor contractions are shown with light gray lines, and the gray circles pinned to the outer legs indicate that these indices are fixed to particular values and hence no longer open. The open tensor legs (on sites i and j) are highlighted in yellow. Note that here we depict the case where the distance $|i-j|$ is 1, 1, and $\sqrt{2}$ in (c), (d), and (e) respectively, but different choices of i and j lead to the same types of tensor contractions. In particular, $|i-j|$ can be up to $\sqrt{2}$ in case (c), $\sqrt{10}$ in case (d), and $3\sqrt{2}$ in case (e).

consider an $L \times L = 10 \times 10$ system in our numerics, both the lower layer and upper layer consist of 5^2 tensors, with the tensors on the edges connected as prescribed by periodic boundary conditions.]

U may also be recast as a large matrix in the full Hilbert space; in this case, the columns of U are the approximate eigenstates. In the tensor network picture, we adopt the convention where the upper indices are the row indices (or contravariant/“ket” indices) and the lower indices are the column indices (or covariant/“bra” indices). Thus, to extract a particular approximate eigenstate from U , one fixes all the lower indices to particular values: the resulting tensor network has only upper open indices and is a projected entangled pair state (PEPS). The indices take the values $\{0, c, d\}$, corresponding to the states $|0\rangle$, $|c\rangle$, and $|d\rangle$, which span the single-site Hilbert space.

Crucially, U is unitary and also conserves particle number and type, i.e., each approximate eigenstate is of a definite number of clean and dirty particles. In our numerics, these conditions are enforced on the individual tensors in the network [boxes in Fig. 1(c)], which we call u . (Note that the tensors are generally distinct, as the system lacks translational invariance; here, u denotes a representative tensor.)

To impose particle number conservation, each u is subject to the condition that the element $u_{\alpha', \beta', \gamma', \delta'}^{\alpha, \beta, \gamma, \delta}$ is nonzero only if $\alpha, \beta, \gamma, \delta = \alpha', \beta', \gamma', \delta'$ up to permutation. That is, the four upper indices of u must contain the same number of both c 's and d 's as the four lower indices. u is also unitary, or more specifically, real and orthogonal—it suffices to have a real ansatz since the Hamiltonian is real. This is imposed by setting $u = e^M$, where M is a real anti-symmetric matrix that contains the actual variational parameters.

The quantum circuit structure of U ensures that calculations involving local operators are local tensor network contractions of finite range, independent of the overall system size. For example, consider the quantity $U^\dagger \mathcal{O}_i U$, where $\mathcal{O}_i$ is a single-site operator acting on site i . On sites away from i , the u and $u^\dagger$ tensors will cancel because of unitarity. Thus, to calculate $U^\dagger \mathcal{O}_i U$, one only needs to consider a 4×4 causal cone of the local operator $\mathcal{O}_i$ as depicted in Fig. 7(b).

APPENDIX B: REDUCED DENSITY MATRIX CALCULATIONS

Here, we describe how we calculate the one- and two-site reduced density matrices from the approximate eigenstates in the optimized U and how we then use the density matrices to obtain the correlation functions and correlation lengths.

Let $|\psi\rangle$ be a (randomly selected) approximate eigenstate. The setup of our ansatz as a unitary quantum circuit allows us to obtain the single-site density matrix $\rho_i = \text{Tr}_{\bar{i}}(|\psi\rangle\langle\psi|)$ and two-site density matrix $\rho_{ij} = \text{Tr}_{\bar{i}\bar{j}}(|\psi\rangle\langle\psi|)$ via causal cone calculations. However, the two-site density matrix trivially decomposes as $\rho_{ij} = \rho_i \otimes \rho_j$ unless sites i and j have overlapping causal cones, which implies that one site must lie in a 6×6 region around the other. Thus, for each site i we consider all 35 other sites j for which nontrivial correlations can be calculated. In Fig. 7 we illustrate the tensor network contractions required to calculate ρ_{ij} in various cases. The calculation of ρ_i

involves a diagram that is the same as Fig. 7(d), but with open legs on only one site i .

Because of the square lattice geometry, the distances considered are (in units of the lattice spacing)

$$|i - j| = 1, \sqrt{2}, 2, \sqrt{5}, 2\sqrt{2}, 3, \sqrt{10}, \sqrt{13}, 3\sqrt{2} \\ \approx 1, 1.41, 2, 2.23, 2.83, 3, 3.16, 3.61, 4.24. \quad (\text{B1})$$

From the single-site reduced density matrices, we can calculate the on-site entanglement entropies $S_i = -\text{Tr}(\rho_i \ln \rho_i)$. From the two-site density matrices, we can calculate $S_{ij} = -\text{Tr}(\rho_{ij} \ln \rho_{ij})$ and thus the quantum mutual information (QMI) between sites. We also calculate the two-point correlation functions, which are defined as

$$C_{\sigma\mu}(i, j) = \langle n_{i,\sigma} n_{j,\mu} \rangle - \langle n_{i,\sigma} \rangle \langle n_{j,\mu} \rangle, \quad (\text{B2})$$

where $\langle \dots \rangle = \langle \psi | \dots | \psi \rangle$, so $\langle n_{i,\sigma} \rangle = \text{Tr}(\rho_i n_\sigma)$ and $\langle n_{i,\sigma} n_{j,\mu} \rangle = \text{Tr}(\rho_{ij} (n_\sigma \otimes n_\mu))$, and where $\sigma, \mu \in \{c, d\}$. We then calculate $C_{\sigma\mu}(i, j)$ at various points and perform an exponential fit $C_{\sigma\mu}(i, j) \propto \exp(-|i - j|/\xi_{\sigma\mu})$ to obtain the correlation length $\xi_{\sigma\mu}$. Note that here $C_{\sigma\mu}$ and hence $\xi_{\sigma\mu}$ corresponds to a specific eigenstate $|\psi\rangle$. We average over eigenstates and disorder realizations to obtain the average quantities $\overline{C_{\sigma\mu}}(\Delta)$ and $\overline{\xi_{\sigma\mu}}(\Delta)$ at a given disorder strength Δ .

APPENDIX C: 6×6 RESULTS

To verify the convergence of the results with increasing system size, we performed a set of calculations for a 6×6 system, analogous to the 10×10 calculations described in the main text. The results, i.e., the 6×6 analog of Fig. 2, are shown in Fig. 8. For the disorder strength used in the experiment, we find that the transition occurs at $f_c(\Delta = 28J) = 0.17$, which is somewhat lower than the value 0.20 obtained from the 10×10 simulations.

The results are similar to those of the 10×10 system, although with fewer data points. We note that our method does not have any significant finite-size effects, but instead

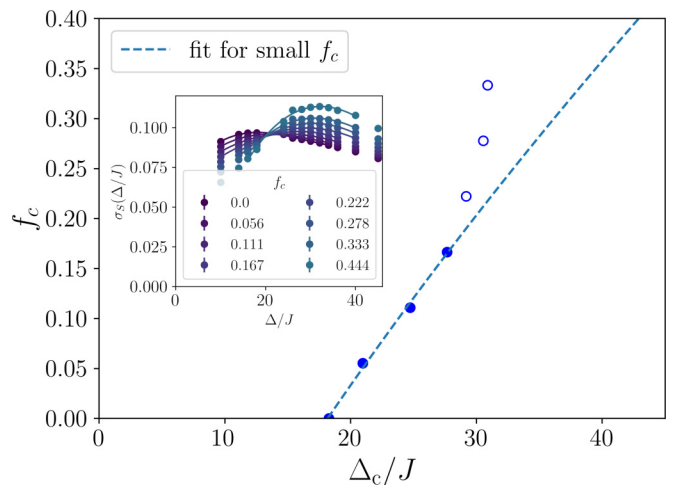


FIG. 8. Clean-fraction resolved phase diagram and (inset) entropy fluctuations, for a 6×6 system, for overall filling fraction $\nu = 0.5$, analogous to Fig. 2 in the main text.

has finite-entanglement effects, as given by the finite range of the causal cones and correspondingly the finite range of correlations our approximation can give rise to. The causal

cones of our approximate local integrals of motion have a finite range, and as a result the unitaries essentially describe the bulk of a two-dimensional infinitely large MBL system.

-
- [1] I. V. Gornyi, A. D. Mirlin, and D. G. Polyakov, Interacting electrons in disordered wires: Anderson localization and low-T transport, *Phys. Rev. Lett.* **95**, 206603 (2005).
 - [2] 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.* **321**, 1126 (2006).
 - [3] M. Žnidarič, T. Prosen, and P. Prelovšek, Many-body localization in the Heisenberg XXZ magnet in a random field, *Phys. Rev. B* **77**, 064426 (2008).
 - [4] A. Pal and D. A. Huse, Many-body localization phase transition, *Phys. Rev. B* **82**, 174411 (2010).
 - [5] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, 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).
 - [6] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe, Many-body localization in a quantum simulator with programmable random disorder, *Nat. Phys.* **12**, 907 (2016).
 - [7] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani *et al.*, Observation of discrete time-crystalline order in a disordered dipolar many-body system, *Nature (London)* **543**, 221 (2017).
 - [8] P. Roushan, C. Neill, J. Tangpanitanon, V. M. Bastidas, A. Megrant, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth *et al.*, Spectroscopic signatures of localization with interacting photons in superconducting qubits, *Science* **358**, 1175 (2017).
 - [9] 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).
 - [10] J. Z. Imbrie, On many-body localization for quantum spin chains, *J. Stat. Phys.* **163**, 998 (2016).
 - [11] J. Šuntajs, J. Bonča, T. Prosen, and L. Vidmar, Quantum chaos challenges many-body localization, *Phys. Rev. E* **102**, 062144 (2020).
 - [12] D. Sels and A. Polkovnikov, Dynamical obstruction to localization in a disordered spin chain, *Phys. Rev. E* **104**, 054105 (2021).
 - [13] J. C. Peacock and D. Sels, Many-body delocalization from embedded thermal inclusion, *Phys. Rev. B* **108**, L020201 (2023).
 - [14] W. D. Roeck and J. Z. Imbrie, Many-body localization: Stability and instability, *Phil. Trans. R. Soc. A* **375**, 20160422 (2017).
 - [15] S. Gopalakrishnan and D. A. Huse, Instability of many-body localized systems as a phase transition in a nonstandard thermodynamic limit, *Phys. Rev. B* **99**, 134305 (2019).
 - [16] J.-Y. Choi, S. Hild, J. Zeiher, P. Schauß, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, Exploring the many-body localization transition in two dimensions, *Science* **352**, 1547 (2016).
 - [17] A. Rubio-Abadal, J.-Y. Choi, J. Zeiher, S. Hollerith, J. Rui, I. Bloch, and C. Gross, Many-body delocalization in the presence of a quantum bath, *Phys. Rev. X* **9**, 041014 (2019).
 - [18] S. Prabhu and E. J. Mueller, Bath mediated decay of density waves in a disordered Bose lattice gas, [arXiv:2104.12648](https://arxiv.org/abs/2104.12648).
 - [19] M. Yan, H.-Y. Hui, M. Rigol, and V. W. Scarola, Equilibration dynamics of strongly interacting bosons in 2D lattices with disorder, *Phys. Rev. Lett.* **119**, 073002 (2017).
 - [20] F. Pollmann, V. Khemani, J. I. Cirac, and S. L. Sondhi, Efficient variational diagonalization of fully many-body localized Hamiltonians, *Phys. Rev. B* **94**, 041116(R) (2016).
 - [21] T. B. Wahl, A. Pal, and S. H. Simon, Efficient representation of fully many-body localized systems using tensor networks, *Phys. Rev. X* **7**, 021018 (2017).
 - [22] T. B. Wahl, F. Venn, and B. Béri, Local integrals of motion detection of localization-protected topological order, *Phys. Rev. B* **105**, 144205 (2022).
 - [23] T. B. Wahl, A. Pal, and S. H. Simon, Signatures of the many-body localized regime in two dimensions, *Nat. Phys.* **15**, 164 (2019).
 - [24] J. Li, A. Chan, and T. B. Wahl, Quantum circuits reproduce the experimental two-dimensional many-body localization transition point, *Phys. Rev. B* **109**, L140202 (2024).
 - [25] F. Venn, T. B. Wahl, and B. Béri, Many-body-localization protection of eigenstate topological order in two dimensions, *Phys. Rev. B* **110**, 165150 (2024).
 - [26] T. B. Wahl, Tensor networks demonstrate the robustness of localization and symmetry-protected topological phases, *Phys. Rev. B* **98**, 054204 (2018).
 - [27] A. Chan and T. B. Wahl, Classification of symmetry-protected topological many-body localized phases in one dimension, *J. Phys.: Condens. Matter* **32**, 305601 (2020).
 - [28] J. Li, A. Chan, and T. B. Wahl, Classification of symmetry-protected topological phases in two-dimensional many-body localized systems, *Phys. Rev. B* **102**, 014205 (2020).
 - [29] J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete, Unifying time evolution and optimization with matrix product states, *Phys. Rev. B* **94**, 165116 (2016).
 - [30] E. V. H. Doggen, I. V. Gornyi, A. D. Mirlin, and D. G. Polyakov, Slow many-body delocalization beyond one dimension, *Phys. Rev. Lett.* **125**, 155701 (2020).
 - [31] P. Crowley and A. Chandran, A constructive theory of the numerically accessible many-body localized to thermal crossover, *SciPost Phys.* **12**, 201 (2022).
 - [32] 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).
 - [33] D. A. Huse, R. Nandkishore, and V. Oganesyan, Phenomenology of fully many-body-localized systems, *Phys. Rev. B* **90**, 174202 (2014).
 - [34] A. Chandran, I. H. Kim, G. Vidal, and D. A. Abanin, Construct-

- ing local integrals of motion in the many-body localized phase, *Phys. Rev. B* **91**, 085425 (2015).
- [35] V. Ros, M. Müller, and A. Scardicchio, Integrals of motion in the many-body localized phase, *Nucl. Phys. B* **891**, 420 (2015).
- [36] S. Inglis and L. Pollet, Accessing many-body localized states through the generalized Gibbs ensemble, *Phys. Rev. Lett.* **117**, 120402 (2016).
- [37] L. Rademaker and M. Ortuño, Explicit local integrals of motion for the many-body localized state, *Phys. Rev. Lett.* **116**, 010404 (2016).
- [38] C. Monthus, Many-body localization: Construction of the emergent local conserved operators via block real-space renormalization, *J. Stat. Mech.* (2016) 033101.
- [39] M. Goihl, M. Gluza, C. Krumnow, and J. Eisert, Construction of exact constants of motion and effective models for many-body localized systems, *Phys. Rev. B* **97**, 134202 (2018).
- [40] A. K. Kulshreshtha, A. Pal, T. B. Wahl, and S. H. Simon, Behavior of l-bits near the many-body localization transition, *Phys. Rev. B* **98**, 184201 (2018).
- [41] A. Chandran, A. Pal, C. R. Laumann, and A. Scardicchio, Many-body localization beyond eigenstates in all dimensions, *Phys. Rev. B* **94**, 144203 (2016).
- [42] D. J. Luitz, N. Laflorencie, and F. Alet, Many-body localization edge in the random-field Heisenberg chain, *Phys. Rev. B* **91**, 081103(R) (2015).
- [43] J. A. Kjäll, J. H. Bardarson, and F. Pollmann, Many-body localization in a disordered quantum Ising chain, *Phys. Rev. Lett.* **113**, 107204 (2014).
- [44] J. Hauschild, J. Unfried, S. Anand, B. Andrews, M. Bintz, U. Borla, S. Divic, M. Drescher, J. Geiger, M. Hefel *et al.*, Tensor network Python (TeNPy) version 1, *SciPost Phys. Codebases* **41** (2024).
- [45] D. Hundertmark, A short introduction to Anderson localization, in *Analysis and Stochastics of Growth Processes and Interface Models* (Oxford Academic, Oxford, 2008), pp. 194–218.
- [46] F. Pietracaprina, V. Ros, and A. Scardicchio, Forward approximation as a mean-field approximation for the Anderson and many-body localization transitions, *Phys. Rev. B* **93**, 054201 (2016).
- [47] A. Scardicchio and T. Thiery, Perturbation theory approaches to Anderson and many-body localization: Some lecture notes, [arXiv:1710.01234](https://arxiv.org/abs/1710.01234).
- [48] G. De Tomasi, S. Bera, J. H. Bardarson, and F. Pollmann, Quantum mutual information as a probe for many-body localization, *Phys. Rev. Lett.* **118**, 016804 (2017).
- [49] B. Villalonga and B. K. Clark, Characterizing the many-body localization transition through correlations, [arXiv:2007.06586](https://arxiv.org/abs/2007.06586).
- [50] A. Morningstar, L. Colmenarez, V. Khemani, D. J. Luitz, and D. A. Huse, Avalanches and many-body resonances in many-body localized systems, *Phys. Rev. B* **105**, 174205 (2022).
- [51] V. Khemani, S. P. Lim, D. N. Sheng, and D. A. Huse, Critical properties of the many-body localization transition, *Phys. Rev. X* **7**, 021013 (2017).
- [52] B. Villalonga and B. K. Clark, Eigenstates hybridize on all length scales at the many-body localization transition, [arXiv:2005.13558](https://arxiv.org/abs/2005.13558).
- [53] H. Ha, A. Morningstar, and D. A. Huse, Many-body resonances in the avalanche instability of many-body localization, *Phys. Rev. Lett.* **130**, 250405 (2023).