Dynamical localization in $\mathbb{Z}_2$ lattice gauge theories

Adam Smith,[1,*] Johannes Knolle,[1,2] Roderich Moessner,[3] and Dmitry L. Kovrizhin[4,5]
[1]*T.C.M. group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*
[2]*Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom*
[3]*Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Straße 38, 01187 Dresden, Germany*
[4]*Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, United Kingdom*
[5]*NRC Kurchatov institute, 1 Kurchatov square, 123182, Moscow, Russia*

(Received 17 March 2018; revised manuscript received 29 May 2018; published 22 June 2018)

We study quantum quenches in two-dimensional lattice gauge theories with fermions coupled to dynamical $\mathbb{Z}_2$ gauge fields. Through the identification of an extensive set of conserved quantities, we propose a generic mechanism of charge localization in the absence of quenched disorder both in the Hamiltonian and in the initial states. We provide diagnostics of this localization through a set of experimentally relevant dynamical measures, entanglement measures, as well as spectral properties of the model. One of the defining features of the models that we study is a binary nature of emergent disorder, related to $\mathbb{Z}_2$ degrees of freedom. This results in a qualitatively different behavior in the strong disorder limit compared to typically studied models of localization. For example, it gives rise to a possibility of a delocalization transition via a mechanism of quantum percolation in dimensions higher than 1D. We highlight the importance of our general phenomenology to questions related to dynamics of defects in Kitaev's toric code, and to quantum quenches in Hubbard models. While the simplest models we consider are effectively noninteracting, we also include interactions leading to many-body localizationlike logarithmic entanglement growth. Finally, we consider effects of interactions that generate dynamics for conserved charges, which gives rise to only transient localization behavior, or quasi-many-body localization.

DOI: 10.1103/PhysRevB.97.245137

I. INTRODUCTION

Gauge theories play a central role in theoretical physics, most famously in the unified description of fundamental particles in the standard model, but also increasingly in the description of condensed matter systems [1–5], where lattice gauge theory (LGT) models often arise as effective descriptions of strongly correlated systems. The celebrated toric code [6] is such an example of a $\mathbb{Z}_2$ lattice gauge theory, which is a prototypical quantum stabilizer code, which also serves as an effective description of the Kitaev honeycomb model with strongly anisotropic couplings [7].

The honeycomb Kitaev model can itself be understood in terms of itinerant Majorana fermions coupled to static $\mathbb{Z}_2$ gauge fields. Other examples include the resonating valence-bond liquid [8,9], slave-particle descriptions of the Hubbard model [10,11], non-Fermi metals [12] and glasses [13], the Falicov-Kimball model [14–17], etc. While models of lattice gauge theories are often difficult to realize in experiment, recent developments in cold atom quantum simulators have opened possibilities in studying these models, see, e.g., the pioneering experiment on cold ion simulations of the Schwinger model [18]. This progress motivates the importance in understanding simple exactly solvable models, which one could use to benchmark experiments and to improve our theoretical understanding of universal behavior of LGTs.

Another field of importance in condensed matter physics is localization (Anderson localization [19] and many-body localization), which has seen recent remarkable developments in theory [20,21] and experiments [22,23]. Localization phenomena provide a set of fundamental concepts about the insulating behavior of itinerant degrees of freedom in the presence of disorder. Remarkably, localization was shown to persist even in presence of interactions and the resulting many-body localized (MBL) phase is a novel exotic dynamical phase of matter [24–26], which is robust to generic perturbations. Many-body localization provides a mechanism for nontrivial relaxation, beyond integrable models, and serves as a counterexample to eigenstate thermalization [27].

Following original ideas of Kagan and Maksimov [28], a number of models for disorder-free localization featuring heavy and light particles have been proposed. In this setup, localization may be induced purely via interactions between the two species [29–32] without any quenched disorder in the Hamiltonian. While numerics suggests localization behavior in these systems, so far, it has been found to be only transient, giving way to ergodic behavior in the long-time limit. This behavior was therefore dubbed quasi-MBL [31]. Another interesting approach is to take quantum analogues of classically-glassy systems, where non-ergodic behavior [33–35] has also been observed. Unfortunately, due to small available system sizes, glassy behavior in these models has not yet been distinguished from that of quasi-MBL.

In previous work [36,37], we have demonstrated, for the first time, a general mechanism for disorder-free localization in a model of fermions with a local $\mathbb{Z}_2$ gauge symmetry. We showed

*as2457@cam.ac.uk

 ©2018 American Physical Society

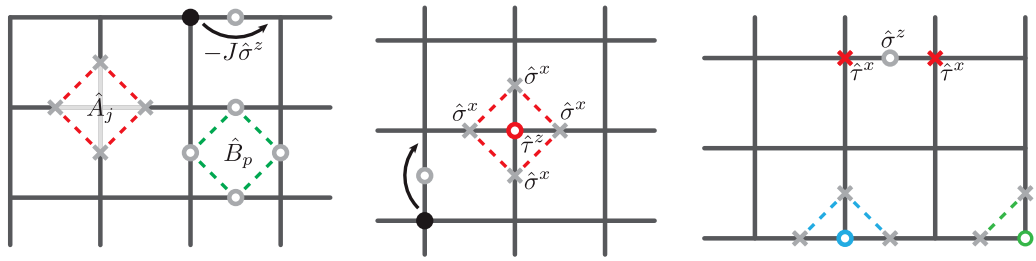

FIG. 1. Schematic picture of the model (1). Left panel shows star $\hat{A}$ and plaquette $\hat{B}$ operators, with $\hat{\sigma}^x$, $\hat{\sigma}^z$ operators denoted by crosses and circles, respectively. Fermion hopping is defined by J and the direction of the spin-1/2 on that bond. Center and right panels show the duality transformation to new spins $\hat{\tau}$. The model can be defined with periodic boundary conditions, as in the center panel. In the case of open boundaries, we define incomplete boundary "stars" as shown in the right panel.

how the localization signatures can be revealed through an exact identification between conserved charges associated with a $\mathbb{Z}_2$ gauge symmetry and an effective binary potential for non-interacting fermions. Moreover, we were able to map fermion or spin correlators to disorder averaged fermionic correlators despite both the Hamiltonian and the initial state having no quenched disorder at all.

In this paper, we extend our theory to a family of $\mathbb{Z}_2$ lattice gauge models of spinless fermions coupled to spins-1/2. These models can be defined in any dimension, and for lattices described by arbitrary graphs, and in particular we focus on two-dimensional square lattices. Here we show that the localization mechanism we discussed in the 1D case applies in a more general context. We use experimentally relevant dynamic probes to diagnose localization [22,23]. In two dimensions, we are able to make a direct connection between our model and Kitaev's toric code model in presence of dynamical charges.

We also analyze in more detail the limit of strong disorder. In the case of $\mathbb{Z}_2$ gauge degrees of freedom, this disorder takes binary values, which gives rise to a phenomenology not found in the typically studied continuous quenched disorder realizations [38]. We find a mechanism for delocalization in 2D related to the phenomenology of quantum percolation [39,40]. Further, we study perturbations that render our models fully interacting. As our preliminary studies in Ref. [37] have indicated, here we demonstrate that in presence of perturbations that do not induce dynamics of conserved charges, the entanglement entropy is characterized by a logarithmic growth, which can be likened to MBL behavior. In addition, we provide an analysis of the effects of perturbations that generate dynamics of conserved charges, which leads to quasi-MBL behavior in the region of parameters that we have explored.

The structure of the paper is the following. In Sec. II, we define the family of models in arbitrary spatial dimension and on arbitrary lattices. We identify conserved charges which reveal the general disorder-free mechanism for localization. We provide details about the transformations, the initial states, and the calculations. We explain the phenomenology of disorder-free localization for the example of 2D Kitaev's toric code and the dynamics of defects therein in Sec. II C. In Sec. III, we present a discussion on the diagnostics of localization behavior. In Secs. III B and III D, we focus on the binary nature of the effective disorder, which leads to qualitative differences in behavior. In Sec. IV, we discuss the effects of integrability-breaking perturbations. In Sec. IV A, we discuss MBL physics which arises in presence of perturbations, while in Sec. IV B, we consider terms that give dynamics to our effective disorder, leading to quasi-MBL behavior. A general discussion and conclusions are presented in Sec. V. We provide in-depth details of all of the numerical methods used in the paper in Appendices.

II. MODEL

We study a family of lattice models with spinless fermions $\hat{f}_i$, which live on the sites of a lattice minimally coupled to spins 1/2 $\hat{\sigma}_{jk}$, positioned on the bonds. These models can be defined on an arbitrary graph, however, in this paper we focus on one-dimensional chains and a two-dimensional square lattice with both open and periodic boundary conditions. We also discuss three-dimensional generalizations, as well as effects of perturbations. The models are described by the Hamiltonian

$$\hat{H} = -\sum_{\langle jk\rangle} J_{jk}\hat{\sigma}^z_{jk}\hat{f}^\dagger_j\hat{f}_k - \sum_j h_j\hat{A}_j, \tag{1}$$

where $\langle jk\rangle$ denotes nearest neighbors, and $\hat{A}_j$ is the *star* operator, which is the product of all spins on the bonds connected to site j, shown for a 2D square lattice in Fig. 1,

$$\hat{A}_j = \prod_{k:\langle jk\rangle}\hat{\sigma}^x_{jk}, \tag{2}$$

and J_{jk} and h_j define coupling strength, and local magnetic field, respectively. In the following, we assume that both of them are position independent. The Hamiltonian possesses an extensive number of conserved quantities (charges) $\hat{q}_i = (-1)^{\hat{n}_i}\hat{A}_i$, where $\hat{n}_i = \hat{f}^\dagger_i\hat{f}_i$. The charges have eigenvalues ± 1 and commute with the Hamiltonian and amongst themselves $[\hat{H},\hat{q}_i] = 0$, and $[\hat{q}_i,\hat{q}_j] = 0$. They can be used to generate local $\mathbb{Z}_2$ gauge transformations under which the Hamiltonian is invariant. Explicitly, these transformations are given by the unitary operators $\hat{U}(\{\theta_i\}) = \prod_i \hat{q}_i^{(1-\theta_i)/2}$, where $\theta_i = \pm 1$, which transform the operators accordingly,

$$\hat{f} \to \theta_i\hat{f}, \quad \hat{\sigma}^z_{ij} \to \theta_i\theta_j\hat{\sigma}^z_{ij}. \tag{3}$$

It is worth noting that our model is an example of an unconstrained $\mathbb{Z}_2$ lattice gauge theory. Explicitly while the Hamiltonian is invariant under the gauge transformation, the Hilbert

space is not. What is typically understood as a gauge theory is constrained to the physical subspace of gauge invariant states by Gauss law [1–3] $\hat{q}_i|\Psi\rangle = |\Psi\rangle$, which we do not impose in our case, cf. the gauge structure of Kitaev honeycomb model [7].

In our previous work [36,37], we studied the Hamiltonian (1) defined on a 1D chain. In this case, the star operators reduce to nearest-neighbor exchange couplings $\hat{A}_j = \hat{\sigma}^x_{j-1,j}\hat{\sigma}^x_{j,j+1}$, and the Hamiltonian (1) assumes the following form:

$$\hat{H}_{1\mathrm{D}} = -J\sum_{\langle jk\rangle}\hat{\sigma}^z_{jk}\hat{f}^\dagger_j\hat{f}_k - h\sum_j \hat{\sigma}^x_{j-1,j}\hat{\sigma}^x_{j,j+1}. \tag{4}$$

Previously, we have shown for this 1D model that conserved charges $\hat{q}_i$ play a role of emergent binary disorder, which gives rise to localization of electron degrees of freedom.

Models described by the class of Hamiltonians (1) appear in the studies of a wide range of systems. Specifically, by imposing a global constraint on the conserved charges such that $\sum_i(\hat{q}_i+1)|\Psi\rangle = \mu|\Psi\rangle$, where $\mu\in\mathbb{Z}$, we recover the Falicov-Kimball model [14–16], where μ/N is the density of localized electrons, see discussion below. Further, in the case of spin-1/2 fermions with the kinetic term described by $\sum_{\alpha=\uparrow,\downarrow,\langle ij\rangle}\hat{\sigma}^z_{ij}\hat{f}^\dagger_{i\alpha}\hat{f}_{j\alpha}$ and imposing the Gauss law constraint, we arrive at the slave-boson description of the Hubbard model [11]. For a recent exposition of other interesting directions where the phenomenology of disorder-free localization described in this paper applies see, e.g., Ref. [41]. Also see Ref. [42], which studies a mechanism similar to ours in the Schwinger model with discrete $\mathbb{Z}_2$ symmetry replaced by a continuous $U(1)$ gauge field.

A. Duality mapping

In our previous work [36,37], we found a mapping that reveals an equivalence between charge configurations $\{q_i = \pm1\}$ and configurations of on-site potentials for the fermions. Here we explain this mapping in greater detail and for the more general class of models described by Eq. (1).

We proceed by a duality transformation of the operators σ, defining spin-1/2 operators τ, which live on the sites of the lattice,

$$\hat{\tau}^z_j = \hat{A}_j, \quad \hat{\tau}^x_j\hat{\tau}^x_k = \hat{\sigma}^z_{jk}, \tag{5}$$

where indices j and k correspond to nearest neighbor sites, see Fig. 1. We have to choose one of the disconnected subspaces of the model. These disconnected subspaces can be enumerated by another set of conserved quantities defined as products of $\hat{\sigma}^z$ along closed loops on the lattice. These conserved quantities can be expressed in terms of plaquette operators, $\hat{B}_p$, defined on the irreducible plaquettes of the lattice (see Fig. 6) and Wilson loop operators $\hat{\Gamma}_n$,

$$\hat{B}_p = \prod_{\text{plaquette } p}\hat{\sigma}^z_{jk}, \quad \hat{\Gamma}_n = \prod_{\langle jk\rangle\in\gamma_n}\hat{\sigma}^z_{jk}, \tag{6}$$

where γ_n is any closed path that winds around a torus and which cannot be written as a product of plaquette operators. For concreteness, we give two examples: in a 1D periodic chain, there is only one such operator which is the loop around the entire system $\prod\hat{\sigma}^z$. This is equivalent to a statement that a number of domain walls modulo 2 is conserved. On a 2D torus we have $\hat{B}_p$ on all square plaquettes of the lattice and the two Wilson loops around the two periodic directions. Importantly, as well as commuting with the Hamiltonian, these operators commute with the generators of the $\mathbb{Z}_2$ gauge symmetry $\hat{q}_i$. The eigenvalues ±1 of these operators label subspaces which are disconnected under gauge transformations.

The duality mapping in Eq. (5) forces the choice $\hat{B}_p = 1$ on all plaquettes and all $\hat{\Gamma}_n = 1$. For a discussion of different plaquette sectors, see Ref. [41]. In all but one dimension with open boundary conditions, we also have a global constraint which is due to the product of all star operators being equal to the identity, i.e., $\prod_{\text{all } i}\hat{A}_i = 1$. On the other side of the duality mapping this amounts to the constraints

$$\prod_{\text{all } i}\hat{\tau}^z_i = 1, \quad \prod_{\text{all } i}\hat{q}_i = (-1)^{\hat{N}_f}, \tag{7}$$

where $\hat{N}_f = \sum_i\hat{n}_i$ is the total fermion number.

In terms of the τ spins, the Hamiltonian assumes the form

$$\hat{H} = -J\sum_{\langle jk\rangle}\hat{\tau}^x_j\hat{\tau}^x_k\hat{f}^\dagger_j\hat{f}_k - h\sum_j\hat{\tau}^z_j. \tag{8}$$

Although this Hamiltonian is equivalent to Eq. (1) only on a restricted Hilbert space, we will not use notation to distinguish between the two. This form is identical to the one exposed for the 1D chain [36,37], but here the nearest-neighbor connectivity can be described by any graph. In this form we can identify local conserved quantities $\hat{q}_j = \hat{\tau}^z_j(-1)^{\hat{n}_j}$ with $\hat{n}_j = \hat{f}^\dagger_j\hat{f}_j$. These *charges* commute with the Hamiltonian and amongst themselves. The charges are precisely those that generate the gauge symmetry identified in the original degrees of freedom.

Finally, by a change of variables $\hat{c}_j = \hat{\tau}^x_j\hat{f}_j$, the Hamiltonian can be written in terms of conserved charges and the spinless fermions $\hat{c}$:

$$\hat{H} = -J\sum_{\langle jk\rangle}\hat{c}^\dagger_j\hat{c}_k + 2h\sum_j\hat{q}_j(\hat{c}^\dagger_j\hat{c}_j - 1/2), \tag{9}$$

where we have used the fact that $\hat{n}_j = \hat{f}^\dagger_j\hat{f}_j = \hat{c}^\dagger_j\hat{c}_j$, since $(\hat{\tau}^x_j)^2 = 1$. The canonical commutation relations $\{\hat{c}^\dagger_j,\hat{c}_k\} = \delta_{jk}$ can be similarly verified. For a given charge configuration—that is in the subspace of fixed $\{q_j\} = \pm1$—Hamiltonian (9) describes a tight-binding model with a binary potential whose sign is set at each site by the value of q_j. Note that we recover the Falicov-Kimball model if we impose the global constraint $\sum_i(\hat{q}_i+1)|\Psi\rangle = \mu|\Psi\rangle$, where μ is an integer that corresponds to the chemical potential for the static auxiliary fermions $\hat{g}_j$, defined via $\hat{q}_i = \hat{g}^\dagger_i\hat{g}_i$.

As well as understanding how the operators transform under the mapping, we must also make an identification between the eigenstates. Let us consider tensor product states of the form $|\Psi\rangle_{\sigma,f} = |S\rangle_\sigma\otimes|\psi\rangle_f$, which we wish to identify with a state $|\Psi\rangle_{\tau,c}$ in the Hilbert space of the τ and c degrees of freedom, and in turn with $|\Psi\rangle_{q,c}$. If for the fermion states we choose the Fock states, i.e., $|\psi\rangle_f = \hat{f}^\dagger_j\cdots\hat{f}^\dagger_l|\text{ vacuum}\rangle$, then these states take the same form for the c fermions, and we will drop the subscript in the following. Without loss of generality, let us consider spins in the z-polarized state $|\uparrow\uparrow\uparrow\cdots\rangle_\sigma$—any other spin state in the sector defined by all $\hat{B}_j = 1$ can be reached via

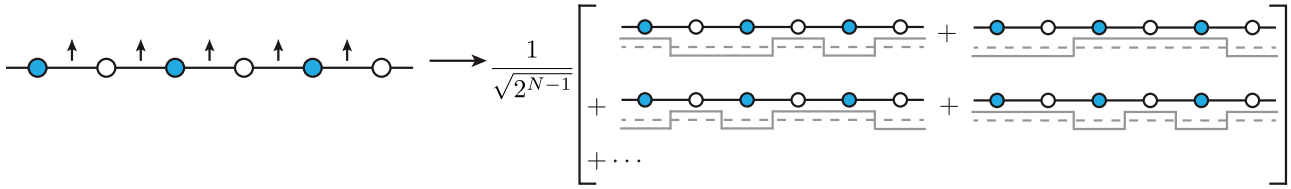

FIG. 2. Schematic picture showing the transformation of the initial state into a dual representation. On the left is an initial state with fermions in a charge density wave (filled sites—blue, empty sites—white). The bond spins are polarized along the z axis. The dual state (right panel) has the fermions in the same configuration, but the wave function is an equal superposition of all charge configurations. In the dual representation, the charge sector leads to exact averaging over the binary potential, which is shown in gray.

application of star operators $\hat{A}_j$. By the duality transformation and the global constraint of Eq. (7), we have that

$$\hat{\tau}_i^x \hat{\tau}_j^x |\uparrow\uparrow\uparrow\cdots\rangle_\sigma = \prod_{\text{all } i} \hat{\tau}_i^z |\uparrow\uparrow\uparrow\cdots\rangle_\sigma = |\uparrow\uparrow\uparrow\cdots\rangle_\sigma, \quad (10)$$

and we make the correspondence between states $|\uparrow\uparrow\uparrow\cdots\rangle_\sigma = \frac{1}{\sqrt{2}}(|\rightarrow\rightarrow\cdots\rangle_\tau + |\leftarrow\leftarrow\cdots\rangle_\tau)$. Therefore we can generally make identifications of the form $|S\rangle_\sigma \otimes |\psi\rangle \propto |S\rangle_\tau \otimes |\psi\rangle$, where $|S\rangle_{\sigma(\tau)}$ has a definite local $z(x)$ component of spin, and $|\psi\rangle$ has a definite local occupation.

Now we can express these tensor product states in terms of conserved charges in place of the τ spins. For a z-polarized state this proceeds as follows:

$$|\uparrow\uparrow\uparrow\cdots\rangle_\sigma \otimes |\psi\rangle = \frac{1}{\sqrt{2^{N-1}}} \sum_{\{\tau_i\}=\uparrow,\downarrow}{}' |\tau_1,\tau_2\cdots\rangle_\tau \otimes |\psi\rangle, \quad (11)$$

where we have identified $|\rightarrow\rangle_\tau = (|\uparrow\rangle_\tau + |\downarrow\rangle_\tau)/\sqrt{2}$, for each τ spin, and the prime indicates that the sum runs over all configurations satisfying constraint (7). Let us consider a single state in this sum $|\tau_1\tau_2\cdots\rangle_\tau \otimes |\psi\rangle$, then since the fermion state is a simple tensor product of site occupation, this can be rewritten as

$$|\tau_1(-1)^{n_1},\tau_2(-1)^{n_2},\cdots\rangle_q \otimes |\psi\rangle. \quad (12)$$

The occupation numbers for the fermion state are fixed and thus only contribute a common sign structure to the charge configuration. Since we sum over all τ configurations, all with a positive weight, this equates to a sum over all charge configurations:

$$|\uparrow\uparrow\cdots\rangle_\sigma \otimes |\psi\rangle = \frac{1}{\sqrt{2^{N-1}}} \sum_{\{q_j\}=\pm1}{}' |q_1,q_2,\cdots,q_N\rangle \otimes |\psi\rangle, \quad (13)$$

where again the primed sum indicates the constraint (7). This transformation of states is shown schematically in Fig. 2. The fact that all of the weights are equal and positive is important for this final form, otherwise there would be a sign structure that depends both on the spin and the fermion configuration. Other spin states in the same spin sector can be accessed through the application of star operators.

B. Emergent disorder and disorder averaging

In the previous section, we showed a transformation to the Hamiltonian (9) which has an effective binary potential, and that the states in the dual configuration are superpositions of the states with a given charge configuration, with the sign of the charges generating a potential for the fermions. Throughout this paper—except briefly in Sec. IIC—we will consider the quenched initial states $|\Psi\rangle = |\uparrow\uparrow\uparrow\cdots\rangle_\sigma \otimes |\psi\rangle$ with z-polarized spins and a selection of fermion Slater determinants. In order to make a connection with the localization problem, the final step is to show that expectation values of observables with respect to these initial states amount to averages over effective disorder.

Let us for concreteness consider a spin expectation value

$$\langle\Psi|\hat{\sigma}_{jk}^z(t)|\Psi\rangle = \frac{1}{2^{N-1}} \sum_{\{s_l\},\{q_m\}=\pm1}{}' \langle\psi|\langle s_1,\cdots| \times e^{i\hat{H}t}\hat{\tau}_j^x\hat{\tau}_k^x e^{-i\hat{H}t}|q_1,\cdots\rangle|\psi\rangle. \quad (14)$$

In order to simplify expressions, we introduce the fermion Hamiltonian

$$\hat{H}(\{q_j\}) = -J\sum_{\langle jk\rangle} \hat{c}_j^\dagger \hat{c}_k + 2h\sum_j q_j(\hat{c}_j^\dagger\hat{c}_j - 1/2). \quad (15)$$

The difference with Eq. (9) is that this Hamiltonian (15) acts only in the fermion subspace and the q_j are no longer operators—the configuration $\{q_j\} = \pm1$ is specified. Equation (14) can then be written as

$$\langle\Psi|\hat{\sigma}_{jk}^z(t)|\Psi\rangle = \frac{1}{2^{N-1}} \sum_{\{q_i\}=\pm1}{}' \langle\psi|e^{i\hat{H}(\{q_i\})t}e^{-i\hat{H}(\{q_i:\bar{q}_j,\bar{q}_k\})t}|\psi\rangle, \quad (16)$$

where $\bar{q}_j$ signifies that the value of charge q_j in the Hamiltonian has been reversed. This reversal of the charge arises from commuting τ operators past the time evolution operator. Note that one can also then remove the second sum over charges because of charge conservation. Similar arguments can be used to show that all correlators of this form reduce to fermion correlators averaged over all charge configurations, which amounts to all disorder configurations in the Hamiltonian (15) [43,44]. The correlators can be efficiently computed using determinants, see Appendix A. It is important to note that, as in Eq. (16), the expressions for the correlators that we obtain are distinct from, e.g., the fermion correlators of a tight-binding model with disorder. For instance, the Green's function

$$\langle\hat{f}_j^\dagger(t)\hat{f}_k(0)\rangle = \frac{1}{2^{N-1}} \sum_{\{q_i\}=\pm1}{}' \langle\psi|e^{i\hat{H}(\{q_i\})t}\hat{c}_j^\dagger e^{-i\hat{H}(\{q_i:\bar{q}_j\})t}\hat{c}_k|\psi\rangle, \quad (17)$$

does not correspond to averaging over disorder configurations for the Green's functions $\langle \hat{c}_j^\dagger(t)\hat{c}_k(0)\rangle$ because of the flipped charges between the forward and backward time evolution. In this respect, the correlators that we obtain are similar to the ones appearing in the x-ray edge problem and the dynamical structure factor for the honeycomb Kitaev model [45,46]. On the other hand, density averages and correlators do equate to simple disorder-averaging without flipped charges.

C. Defect attachment in the toric code model

Before moving on to study the physics of localization in our model, let us first consider Hamiltonian (1) defined on a 2D square lattice with periodic boundary conditions. This becomes equivalent to Kitaev's toric code—with plaquette dynamics frozen—coupled to spinless fermions. Note that one way of introducing dynamical defects into the toric code is by adding transverse field terms $\sum_{\langle jk\rangle}\hat{\sigma}^z_{jk}$. Here we briefly outline the dynamics induced by our coupling to fermions.

For our discussion of localization, we study the initial states of spins polarized along the z axis. However, let us for a moment focus on spin states describing the ground state of the toric code, that is,

$$\hat{A}_j|S_0\rangle = |S_0\rangle, \quad \hat{B}_p|S_0\rangle = |S_0\rangle. \tag{18}$$

We note that our choice of duality transformation is consistent with this ground state, and it fixes Wilson loop operators to be $\hat{\Gamma}_1 = \hat{\Gamma}_2 = 1$, thus it uniquely chooses one of the four degenerate ground states (for $J = 0$) of the toric code. To access other ground state sectors we can modify the duality transformation (5) by defining a vertical and horizontal line (going through bonds) over which the implicit definition of $\hat{\tau}^x$, picks up a sign $\Gamma_1,\Gamma_2 = \pm 1$, respectively. More explicitly, one can define

$$\hat{\tau}^x_j\hat{\tau}^x_k = (\Gamma_1)^{\delta^1_{jk}}(\Gamma_2)^{\delta^2_{jk}}\hat{\sigma}^z_{jk}, \tag{19}$$

where $\delta^{1(2)}_{jk}$ is 1 when the bond $\langle jk\rangle$ crosses a vertical(horizontal) reference line, and 0 otherwise. Note that this choice changes the action of the Wilson loop operators but not $\hat{B}_p$, since any plaquette crosses any line an even number of times.

Having chosen the initial spin configuration, we can consider the coupling to fermions. For a simple tensor product state $|S_0\rangle \otimes |\psi\rangle$, this maps to

$$|\uparrow\uparrow\cdots\rangle_\tau \otimes |\psi\rangle = |(-1)^{\hat{n}_1},(-1)^{\hat{n}_2},\cdots\rangle_q \otimes |\psi\rangle, \tag{20}$$

that is, for an initial fermion state of definite local occupation, the charge configuration is uniquely specified by the parities of fermion occupation numbers on each site. The Hamiltonian then takes a simple form,

$$\hat{H}_{\rm TC} = -J\sum_{\langle jk\rangle}\hat{c}_j^\dagger\hat{c}_k + 2h\sum_j q_j(\hat{c}_j^\dagger\hat{c}_j - 1/2), \tag{21}$$

where in contrast to Eq. (9), the potential given by q_j is fixed and equal to -1 if there is a fermion on this site in the initial state, and $+1$ if the site is empty. If we consider the limit $h \gg J$, then the fermions lie at the bottom of large potential wells, and fermion hopping is suppressed, and we recover the static toric code. The form of the conserved charges is essentially a statement that defects in the toric code are attached to fermions (or holes).

Let us now consider excitations of this model in the limit of $h \gg J$. Addition of a star defect on a site amounts to flipping a τ-spin and corresponds to changing the sign of the potential on the same site. This defect is then free to move in a restricted geometry on the lattice that is determined by the sites occupied by fermions, since the defect is attached to a fermion (hole). This geometry corresponds to a connected region of the lattice which has the opposite fermion parity to the fermion/hole attached to the defect. This can be understood as site percolation problem for the defects, and we will encounter it again in Sec. III D. Importantly, on a square lattice, the percolation threshold is $p_c \approx 0.5927$, which means that for fermions at half-filling and in a random configuration, the defects are localized.

Let us now return to the discussion of the plaquette operators. We can include the latter in our model and through the symmetries of the model do exactly the same procedure of attaching defects to fermions. If we consider a dual lattice, which is a square lattice with sites at the centres of plaquettes, then the plaquette operators become star operators and vice versa. We can then add a second fermion species on this dual lattice which will be attached to plaquette defects. If we denote fermions attached to star defects $\hat{a}$ and those attached to plaquette defects $\hat{b}$, then the toric code model with fermion attachment reads

$$\hat{H}_{\rm TC} = -h_A\sum_s \hat{A}_s - h_B\sum_p \hat{B}_p - J_A\sum_{\langle ij\rangle_s}\hat{\sigma}^z_{j,k}\hat{a}_i^\dagger\hat{a}_j - J_B\sum_{\langle ij\rangle_p}\hat{\sigma}^x_{j,k}\hat{b}_i^\dagger\hat{b}_j, \tag{22}$$

where $\langle ij\rangle_s$ denotes nearest neighbors on the original lattice, and $\langle ij\rangle_p$ denotes those on the dual lattice. The conserved charges are $\hat{q}^A_j = \hat{A}_j(-1)^{\hat{a}_j^\dagger\hat{a}_j}$ and $\hat{q}^B_j = \hat{B}_j(-1)^{\hat{b}_j^\dagger\hat{b}_j}$, respectively. We now have a full toric code with defects attached to fermions. We are then able to control the dynamical behavior of the defects by choosing the initial configuration and occupation numbers for the fermions.

III. LOCALIZATION

In this section, we discuss the localization behavior of the model defined by Eq. (1) in the case of a 1D chain and a 2D square lattice. Since the model can be mapped to free fermions, we can calculate correlators using determinants as explained in Appendix A. This approach allows us to study systems with $\sim 10^2$–10^3 sites. For entanglement entropy calculations, we require the full density matrix and resort to exact diagonalization in 1D for up to $N = 12$ sites. To calculate the density of states, we use the kernel polynomial method [47], see Appendix C, which can be used for systems of order 10^5–10^6 sites. Localization lengths are computed using a standard transfer matrix approach [48] described in Appendix D.

A. Dynamical localization in 1D

The quench problem that we study here has initial states with bond spins polarized along the z axis $|\uparrow\uparrow\uparrow\cdots\rangle$, and with

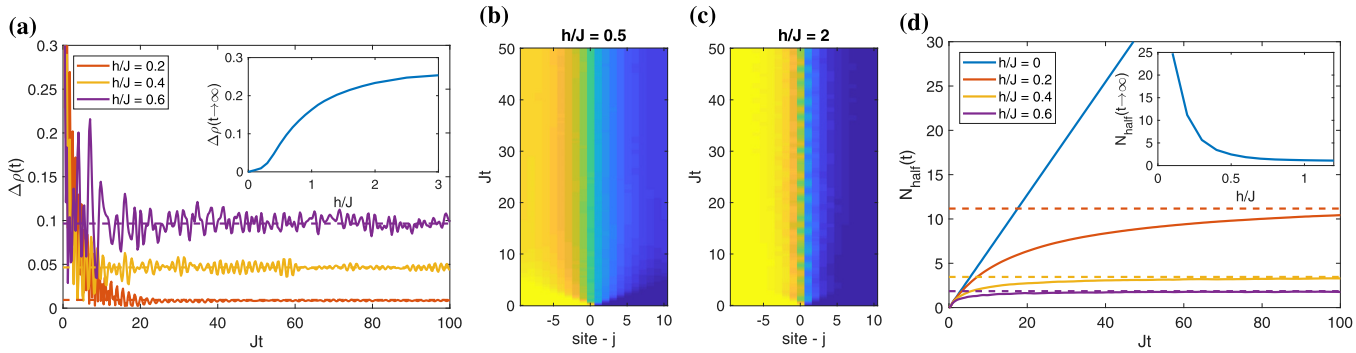

FIG. 3. Results for the time-evolution of the fermion subsystem in 1D. (a) Persistence of a charge density wave measured by the density imbalance between neighboring sites, $\Delta\rho$, see Eq. (23), for $N = 200$ sites with periodic boundary conditions. The inset shows results in the long-time limit, $Jt = 10^9$, as a function of effective disorder strength h/J. (b) and (c) Spreading of a domain wall (in the initial configuration) for $h/J = 0.5$ and 2, respectively, for the system with $N = 20$ sites and open boundary conditions. Local density is shown (filled—yellow, empty—blue). (d) Time evolution of particle occupation numbers on the right-half of the system N_{half} from an initial domain wall configuration for $N = 200$ sites. Inset shows results for the long time limit as a function of h/J. All calculations are performed using the determinant method of Appendix A.

fermions in one of the following Slater determinant states. (i) *Charge density wave* described by fermions in a Fock state with occupation numbers $|\cdots 1010\cdots\rangle$. We will probe the memory of this initial state via the nearest-neighbor density imbalance

$$\Delta\rho(t) = \frac{1}{\widetilde{N}} \sum_j |\langle\Psi|\hat{n}_j(t) - \hat{n}_{j+1}(t)|\Psi\rangle|, \tag{23}$$

where $\widetilde{N} = N - 1, N$, for open and periodic boundary conditions, respectively. This measure was used, e.g., to identify the MBL transition in cold atom experiments, see Ref. [23].

(ii) *Domain wall* configuration with the left half of the chain filled and the right half empty $|\cdots 111000\cdots\rangle$. In order to quantify localization in this case, we measure the total number of particles in the right half of the system (which is empty in the initial state), which tells us how many particles make it across the domain wall,

$$N_{\text{half}}(t) = \sum_{j\in \text{ right half}} \langle\Psi|\hat{n}_j(t)|\Psi\rangle. \tag{24}$$

This observable, as well as the long-time fermion density distribution, reveal the extent to which the fermions are localized. A similar measurement was used to identify the MBL transition in 2D, see the cold atom experiments of Ref. [22], and in theoretical work as a dynamical measure of localization, see Ref. [49].

Let us first consider the charge density wave initial state, where we measure the density imbalance $\Delta\rho(t)$, see Fig. 3(a). While the latter decreases as a function of time from its initial value of 1, at long times it approaches a nonzero value for all $h \neq 0$. Furthermore, the asymptotic value $\Delta\rho(t \to \infty)$ increases monotonically with h/J, as shown in the inset. This shows the memory of the initial state, which is preserved in breaking of ergodicity due to emergent disorder. We also note that fluctuation in the asymptotic value of the density imbalance increase both in amplitude and longevity with increasing h. These fluctuations can be linked to the behavior of the density of states shown in Fig. 5(a), where we see that the DOS shows multiple spikes at the band edges that lead to resonances.

Next, let us consider the domain wall configuration shown in Figs. 3(b)–3(d). For $h/J < 1$, we observe initial linear spreading of the domain wall at short times. At longer times, this linear spreading halts, and the density quickly approaches its limiting value with exponential tails set by the single-particle localization length [36], see Fig. 7(b). For $h/J > 1$, we see a similar phenomenology except the halting is much more abrupt and the domain wall spreads over only a few sites, see Fig. 3(c). The persistence of the domain wall can be most clearly quantified by $N_{\text{half}}(t)$: the number of particles that make it into the initially empty half of the system, as shown in Fig. 3(d). We see that for $h \neq 0$ its behavior deviates from linear growth, which we observe for $h = 0$. The number of particles, which makes it across the domain wall, is bounded showing that there is a remaining imbalance between the two halves of the system, and thus the memory of the initial domain wall remains at arbitrary long times.

B. Strong disorder limit

While we find localization for all values of h, we also observe a qualitative change in behavior for $h/J > 1$, which can be traced back to the binary nature of the effective disorder. See, for example, Refs. [38,39] for further discussion of the differences between binary and continuous disorder. In our model, this difference is most evident in the bipartite von Neumann entanglement entropy, shown in Fig. 4, where we observe a plateau with area-law scaling. For both the CDW and the domain wall initial states, we also find that dynamical observables show larger amplitude and longer lived oscillations with increasing h/J. This can be seen as fluctuations in $\Delta\rho$ in Fig. 3(a) and side-to-side fluctuations near the domain wall in Fig. 3(c).

To understand this behavior, let us consider the strong disorder limit, $h/J \gg 1$. Here one can think of the 1D chain as a collection of finite-length disconnected chains, with the distribution of lengths, l, given by $(1/2)^l$, as shown in the inset of Fig. 5(b). In each isolated chain of length l, we then have l single-particle wave functions and energy levels E. In this limit, we can separate the Hamiltonian into $\hat{H}_h$ and $\hat{H}_J$ given

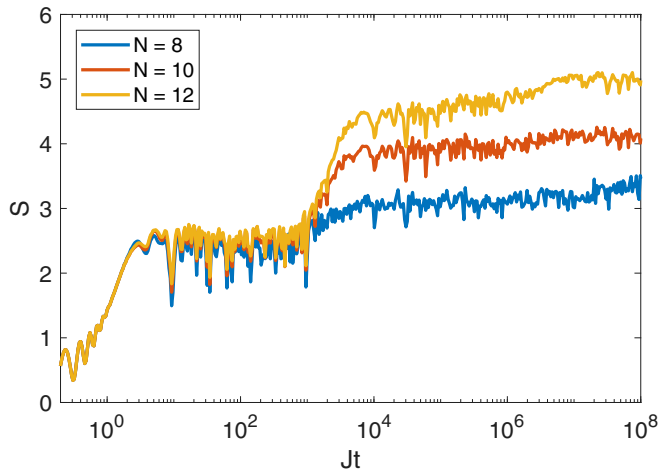

FIG. 4. Time evolution of the bipartite von Neumann entanglement entropy in the 1D case. We start from a charge density wave initial state in a system with open boundary conditions and $h/J = 10$. We partition the system into two halves along the central bond and results are computed using exact diagonalization.

by

$$\hat{H}_h = 2h \sum_i q_i(\hat{c}_i^\dagger \hat{c}_i - 1/2) - J \sum_{\langle ij\rangle: q_i = q_j} (\hat{c}_i^\dagger \hat{c}_j + \text{H.c.}),$$
$$\hat{H}_J = J \sum_{\langle ij\rangle: q_i = -q_j} (\hat{c}_i^\dagger \hat{c}_j + \text{H.c.}). \tag{25}$$

Here, the sums are over nearest neighbors satisfying the condition on the relative sign of q. Note that we omit an overall energy shift $h \sum_i q_i$ in the Hamiltonian, which is defined by a charge configuration. This shift does not affect the results since there is no matrix element between different charge sectors. The Hamiltonian $\hat{H}_h$ describes disconnected uniform tight-binding chains, and $\hat{H}_J$ corresponds to a hopping between these chains.

The DOS of the Hamiltonian $\hat{H}_h$ can be constructed using an ensemble of the energy levels for disconnected chains weighted by their probability distribution using the following equation:

$$g(\omega) \propto -\frac{1}{\pi} \Im m \lim_{\delta \to 0} \sum_{l=1}^{\infty} \sum_{E_l} \frac{(1/2)^l}{\omega - E_l + i\delta}, \tag{26}$$

where E_l denote single-particle eigenvalues of the tight-binding Hamiltonian for a uniform chain of length l. In order to obtain the DOS numerically, we introduce a cutoff on the sum over l and choose a finite broadening $\delta = 0.0015$. This form of the Hamiltonian and the corresponding DOS reveals two main features. First, for $h/J > 1$, we have a gap in the spectrum which splits into two sub-bands of bandwidth $4J$, centered at $\pm 2h$, see inset of Fig. 5(a). Note that in the Falicov-Kimball model this corresponds to the Mott phase [14]. Second, the DOS is characterized by a set of discrete peaks. In Fig. 5(b), we compare the exact DOS centered around one of these sub-bands at $E = 2h$ for a large but finite system with large $h \gg J$, and the DOS constructed from Eq. (26), which shows good agreement.

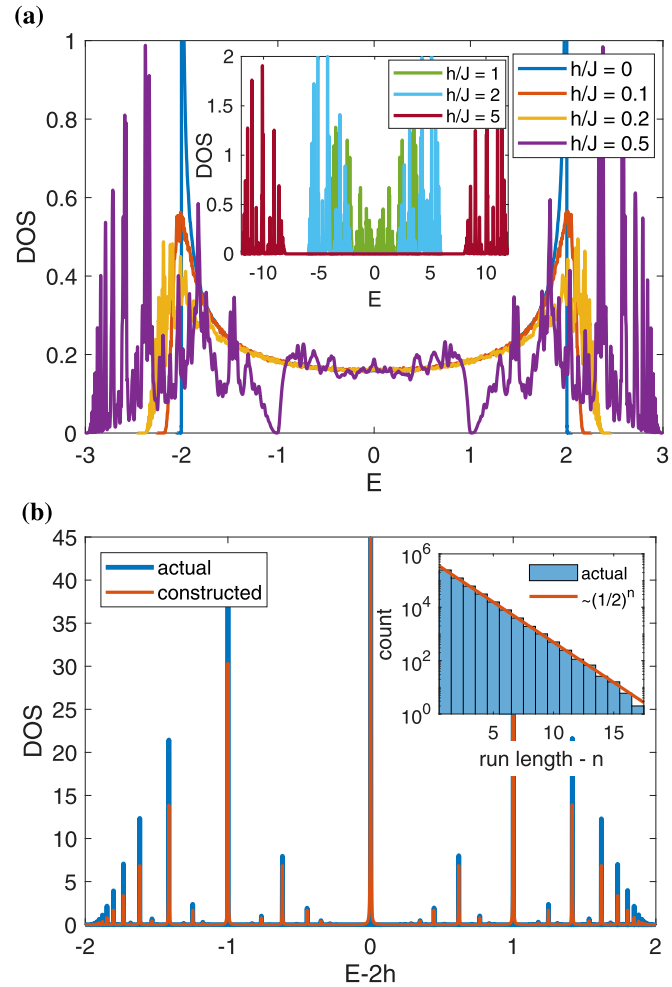

FIG. 5. Density of states for the 1D chain (15). (a) DOS for different values of h/J. Inset shows the DOS for values of $h > J$ (where there is a gap in the DOS). (b) DOS for a very large value of $h/J = 500$. The energy is offset by $2h$ and we focus on one of the two sub-bands that form for large h. The DOS is computed using the kernel polynomial method (see Appendix C), shown in blue. We compare this with the DOS constructed using Eq. (26), shown in red. Inset in (b) shows a comparison of the observed distribution of chain lengths with the corresponding distribution in the thermodynamic limit.

The fact that the DOS splits up into a discrete set, of which only a few carry the majority of the spectral weight, explains the observed fluctuations in our localization diagnostics together with the area-law plateau in the entanglement entropy. These features can be attributed to resonant processes between these few discrete states. Figure 5 shows that a similar structure persists, to some extent, below $h/J = 1$. The effect of $\hat{H}_J$ on the DOS appears at second order in perturbation theory. This gives rise to the broadening of the spectrum, and provides a timescale $\sim(h/J)^2$ which sets the lifetime of the area-law plateau and of the fluctuations in the fermion density. The long-time area-law appears because the spin subsystem relaxation time is given by this timescale $\sim(h/J)^2$ arising from resonant processes. On this timescale the fermion subsystem explores the global structure of the charge distribution leading to a saturation of the entanglement between the charges and the

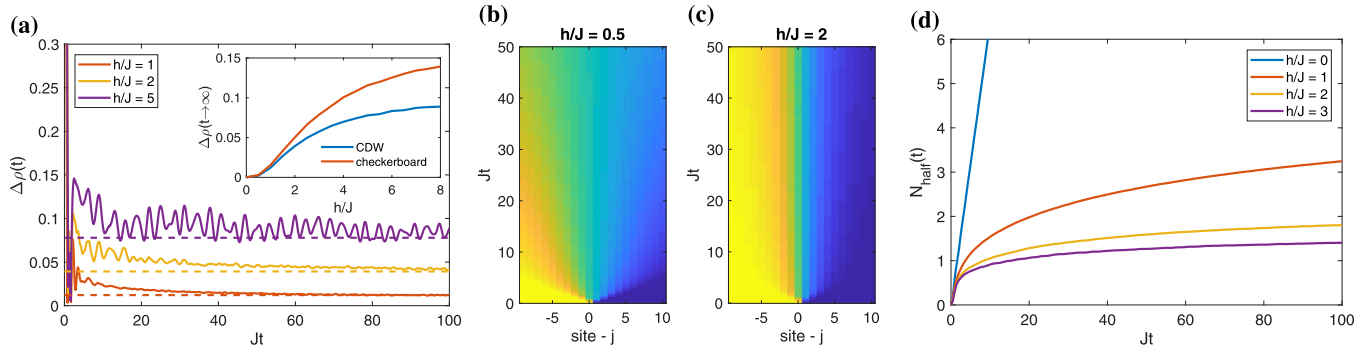

FIG. 6. Time evolution of the fermion subsystem in 2D. (a) Density imbalance $\Delta\rho$ measured along a slice through the center of the system with the initial state described by a charge density wave, see text. Inset shows the long-time limit for the charge density wave, and checkerboard initial states. (b) and (c) Spreading of the domain wall for $h/J = 0.5$ and $h/J = 2$, respectively, measured along the slice through the center of the system. (d) Number of particles N_{half} along the center in the initially empty half of the system. In (a)–(c), we use a square lattice with $N = 32 \times 32$ sites and in (d) $N = 30 \times 50$. Results are computed using the determinant method of Appendix A.

fermions. Differences between binary and continuous disorder are discussed in, e.g., Ref. [38].

C. Dynamical localization in 2D

In two dimensions, we will consider the Hamiltonian (1) on a square lattice, see Fig. 1. As in 1D, we consider initial states with spins polarized along the z axis. We study the initial states of fermions in one of the three following configurations: (i) *charge density wave* with alternating occupation along one of the directions of the lattices and uniform occupation along the other (stripes); (ii) *checkerboard pattern* with alternating occupation along both directions of the lattice; and (iii) *domain wall* configuration with one half of the system filled, and the other empty, such as studied in cold-atom experiments, see Ref. [22].

For all diagnostics we consider a cut through the system, e.g., perpendicular to the domain wall. As in 1D, we find that the density imbalance saturates at a nonzero value at long times, see Fig. 6(a). However, in contrast to 1D, the localization length is larger in 2D (for the same h/J) leading to smaller long-time values for $\Delta\rho$ and larger values for N_{half}. Furthermore, for the values of h/J shown in Fig. 6, which are much larger than those presented for the 1D case, the amplitude of the fluctuations is much smaller. In other words, the extra dimension produces a damping effect on these fluctuations because of the much smoother single-particle DOS, even for $h/J > 1$, see Fig. 7(a). We also find that for the checkerboard initial state, the density imbalance persists more than for the charge density wave, as shown in the inset. This can be understood by considering the action of the Hamiltonian on this initial state. In contrast to the CDW, in the checkerboard case, delocalization of fermions appears at higher order in J/h, thus constraining the fermion dynamics in the checkerboard initial state compared to a charge density wave. Comparison of the corresponding 1D and 2D results shows that the remaining imbalance is generally much smaller in 2D than in 1D, which is due to the fact that localization lengths are much larger in 2D, see Fig. 7(b).

Starting from the domain wall initial states, we can again see a linear initial spreading, which is halted due to the effective disorder, see Figs. 6(b) and 6(c). In this case, we do not find long-lived oscillations for $h/J > 1$. Our results clearly show that the localization length is much larger in

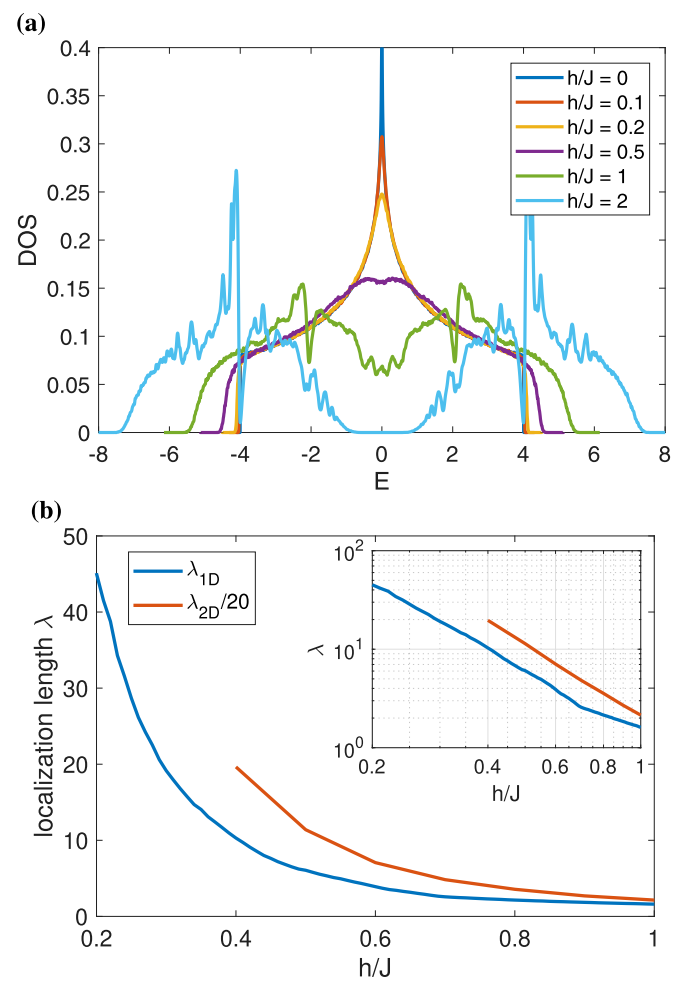

FIG. 7. (a) Single-particle density of states for a 2D square lattice for different h/J. The DOS is computed using the kernel polynomial method, see Appendix C. (b) Localization length in 1D and 2D. In 1D, the localization length is computed using the spectral formula (27) as explained in the main text. In 2D, the localization length is scaled by a factor of 20 and is computed by the transfer matrix method on a strip of width 100 sites and length 250 000 sites.

2D compared to 1D, which can be seen from the dynamics of domain wall spreading (on a larger length scale) for the same disorder strength as in 1D. We can use again N_{half} to quantify the dynamics of domain wall spreading. This observable approaches a finite value in the localized case, see Fig. 6(d).

If we compare the DOS for 2D with that of 1D, we notice some important similarities and differences. First, we see a gap opening as in 1D for large values of h. Whereas in the 1D case this gap appears at $h = J$, the spectrum is still gapless for h order of J in 2D. There is also an increase in the bandwidth, both of the total DOS and of the individual subbands that develop in the large h limit, owing to the extra dimension. More importantly, we find that subbands remain much smoother than in 1D for a much wider range of effective disorder strength.

In Fig. 7, we show the dependence of the maximum localization length on h/J in 1D and 2D. The localization length is the characteristic length scale of the exponential tails of the single-particle wave functions defined via $e^{-j/\lambda}$. In the 1D case, the results are obtained using the spectral formula [48,50]

$$\frac{1}{\lambda} = \min_E \int_{-\infty}^{\infty} g(x) \ln |E - x| \, \mathrm{d}x, \tag{27}$$

where $g(x)$ is the DOS calculated via the kernel polynomial method (Appendix C). In the 2D case, we used the transfer matrix method [48], see Appendix D. The 2D results are rescaled by a factor of 20, which demonstrates an order of magnitude difference in localization lengths in 1D and 2D. However, the localization length as a function of disorder strength shows similar power law in 1D as in 2D.

D. Delocalization in 2D and quantum percolation

It is known that all single-particle states in 1D and 2D are localized in the presence of disorder. However, it is possible to have delocalized states and even a mobility edge separating localized and delocalized states in case of correlated disorder. The famous example in 1D is the Aubry-Andre model, which has a periodic potential incommensurate with the lattice. Another example is when time-reversal symmetry is broken, for example, by a magnetic field.

Due to the binary nature of the disorder potential in our model, we can get delocalized states in 2D without the need of correlated disorder. This can be achieved by biasing the distribution of charges q with a probability p, such that $q = \pm 1$ with probability p and $1 - p$, respectively. Alternatively, one could impose a stricter global constraint $N^{-1} \sum_i (\hat{q}_i + 1)|\Psi\rangle = p|\Psi\rangle$. In the large h limit of strong disorder, we arrive at a quantum site percolation problem [39]. The lattice then is decomposed into parts with sites sitting at the top or bottom of the binary potential, and to the lowest order only hopping between sites with the same sign of potential is allowed. This can be appreciated by considering Eq. (25) on a 2D lattice. In $\hat{H}_h$, there will only be hopping terms between neighboring sites with the same values of q, see Fig. 8. Since the threshold in the classical site percolation problem is $p_c \approx 0.5927$, this is consistent with having localized wave functions for all h for our typical setting of $p = 1/2$. However, if we set $p > p_c$, or $1 - p > p_c$, then we should expect percolation in our model

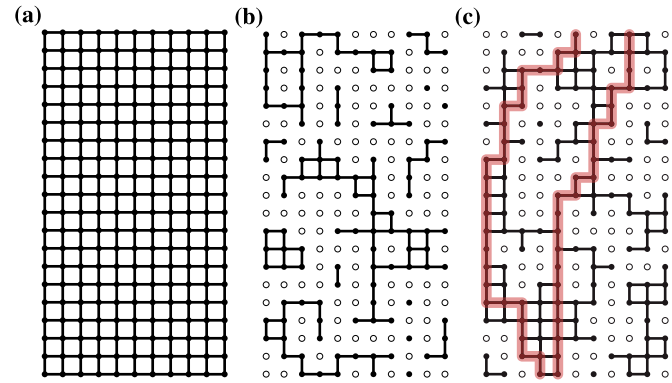

FIG. 8. Schematic picture of the site percolation problem. (a) A fully connected lattice on which fermions can hop in the limit $h/J = 0$. [(b) and (c)] When $h/J \gg 1$, the fermions are constrained to hopping between sites with the same effective potential (filled sites and bonds) and the other sites become inaccessible (open circles) leading to a quantum site percolation problem. (b) Connected sites for the bias $p = 0.5$ showing that the absence of paths (in the thermodynamic limit), which connect opposite sides of the lattice. (c) The same for the bias $p = 0.7$, which is above the percolation threshold $p_c \approx 0.5927$, showing in red connected paths across the system.

for large h. Note that if we get percolation, say at the top of the potential, then we necessarily localize the modes at the bottom of the potential more, and vice versa.

In order to understand the effect of percolation, we study the time evolution from a domain wall initial state with changing system size and bias p, as shown in Fig. 9(a). We plot $N_{\text{half}}(t \to \infty)/L$, where L is the linear dimension of a square lattice with $N = L \times L$ sites. If particles are localized then we would expect $N_{\text{half}}(t \to \infty)/L$ to tend to zero as L is increased—i.e., N_{half} is finite and independent of L. Whereas, if the particles are delocalized, we should find that a finite proportion of the particles makes it across the domain wall, $N_{\text{half}}(t \to \infty)/L \to$ constant. In Fig. 9(a), we show the extrapolation from finite size scaling, which agrees with this expected behavior in the two limits. Furthermore, the change in behavior is observed to happen around the critical percolation threshold $p = p_c$.

We can also understand this percolation behavior by looking at the DOS and the energy-resolved localization length as a function of p, as shown in Fig. 9(b). As p is increased past the critical point, we see a clear asymmetry with respect to energy in these results. In the DOS, one of the sub-bands becomes similar to that of the large h limit of $p = 1/2$ problem. The other sub-band becomes much smoother and similar to the DOS for a clean 2D system. Furthermore, we respectively see a decrease and increase in the localization length in these two sub-bands, see inset, which is consistent with percolation of the fermions at a positive potential and the localization of those at a negative potential.

Because we are studying a quantum model, it is not clear that there should be a direct correspondence with site percolation. In particular, given a path through the system, as in Fig. 8, we would generally expect quantum fluctuations to lead to backscattering, which may hinder conductance. References [17,40] showed, using large scale numerics with the kernel polynomial methods, that the quantum percolation

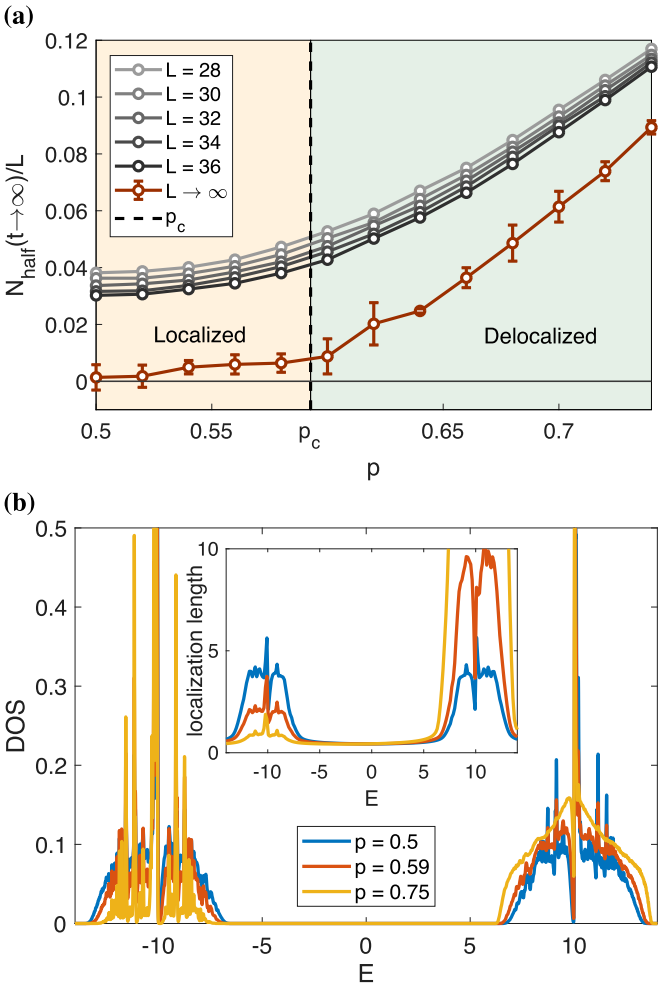

FIG. 9. (a) Total number of particles making it across a domain wall N_{half} for various values of the charge distribution bias, p. The figure shows the rescaled value N_{half}/L, and the extrapolated limit $L \to \infty$, where L is the linear dimension of the system. Error bars correspond to two standard deviations in the linear fit from finite size scaling. Note that $p_c \approx 0.5927$ corresponds to the classical site percolation threshold for a square lattice. (b) Density of states for different values of the bias p and $h/J = 20$. Inset shows the corresponding energy-resolved localization length. The results shown in (a) were obtained using the determinant method (Appendix A). We used the KPM (Appendix C) for the DOS in (b) and the transfer matrix approach for the localization length (Appendix D) for a system of $150 \times 250\,000$ sites.

threshold $p_c^q < 1$, and we necessarily have $p_c \leqslant p_c^q$. Furthermore, studies on the Bethe lattice seem to show that the quantum site percolation threshold agrees with the classical threshold [39], and our results are also consistent with the classical site percolation on a square lattice. However, this point is far from having been settled and Ref. [40] observes that $p_c < p_c^q$ by doing a careful analysis using local density of states for much larger systems. In particular, they find that $p_c^q > 0.65$, which we do not see in our results. However, because of the modest system sizes used in our calculations, our results may still show finite-size effects.

E. Localization in 3D

For a 3D cubic lattice, it is generally believed that $p_c < p_c^q < 1$, though still $p_c^q < 0.5$ [40]. Therefore a 3D case offers an interesting setting for studying localization. In 3D, there is a critical disorder strength needed to achieve localization. However, in the large h/J limit, we would expect delocalized states for all values of the bias probability p. One can expect delocalized states for both low and high h/J but localized states for intermediate values.

IV. LOCALIZATION IN PRESENCE OF INTERACTIONS

Our model can be modified in a variety of ways to include interactions. Here we consider a subset of such extensions, focusing on the 1D case because of numerical limitations. Terms which can be added to the Hamiltonian (1) fall into two classes depending on whether they give dynamics to conserved charges. Note that these terms in general lead to interactions in contrast to Hamiltonian (9). In the presence of interactions, it is not possible to use determinant methods, and instead we have to resort to exact diagonalization and on Krylov subspace methods to calculate the time evolution, see Appendix B.

A. Conserved charges and many-body localization

Those terms that commute with the charges include fermion density-density interactions and longitudinal field terms:

$$\Delta \sum_{\langle jk\rangle} \hat{n}_j \hat{n}_k, \quad \text{and} \quad B_x \sum_j \hat{\sigma}^x_{jk}, \tag{28}$$

where $\hat{n}_j = \hat{f}_j^\dagger \hat{f}_j$.

Commutation relations between these terms and the charges are most easily seen when they are expressed in terms of the original degrees of freedom, $\hat{q}_j = \hat{\sigma}^x_{j-1,j}\hat{\sigma}^x_{j,j+1}(-1)^{\hat{n}_j}$. The charges clearly do not commute with any local terms containing $\hat{\sigma}^z_j, \hat{\sigma}^y_j$, but do commute with a longitudinal magnetic field. Electron density-density interactions retain their form under the transformation from f to c fermions, and the density operator also commutes with the fermion parity operator appearing in the expression for conserved charges.

First, we consider adding nearest-neighbor density-density interactions between fermions. Up to constant terms, the model can be written as

$$\hat{H} = -J \sum_{\langle jk\rangle} \hat{\sigma}^z_{jk} \hat{f}_j^\dagger \hat{f}_k - h \sum_j \hat{\sigma}^x_{j-1,j}\hat{\sigma}^x_{j,j+1} + \Delta \sum_j (2\hat{n}_j - 1)(2\hat{n}_{j+1} - 1). \tag{29}$$

Under transformation to c-fermions and charges, the first two terms transform as before and the fermion interactions remain invariant since $\hat{f}_j^\dagger \hat{f}_j \equiv \hat{c}_j^\dagger \hat{c}_j$. We can then use a Jordan-Wigner transformation, $\hat{S}_j^+ = \hat{c}_j^\dagger (-1)^{\sum_{l<j}\hat{n}_l}$, and $\hat{S}_j^z = \hat{n}_j = \frac{1}{2}$ to cast the Hamiltonian in the following form:

$$\hat{H}_{\text{XXZ}} = -J \sum_j (\hat{S}_j^+ \hat{S}_{j+1}^- + \hat{S}_j^- \hat{S}_{j+1}^+) + 4\Delta \sum_j \hat{S}_j^z \hat{S}_{j+1}^z + 2h \sum_j \hat{q}_j \hat{S}_j^z, \tag{30}$$

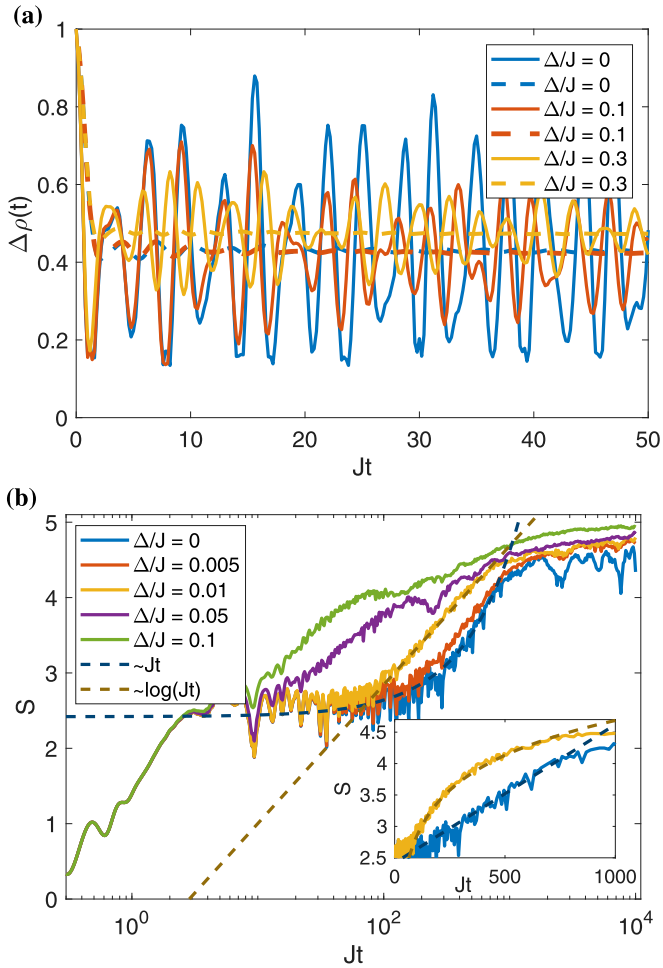

FIG. 10. Quantum quench from an initial charge density wave state, $h/J = 20$. (a) Density imbalance $\Delta\rho(t) \propto \sum_j |\langle 0|\hat{n}_j(t) - \hat{n}_{j+1}(t)|0\rangle|$ and the time-averaged value of $\frac{1}{t}\int_0^t \mathrm{d}\tau\, \Delta\rho(\tau)$ (dashed lines) after the same quench. (b) Von Neumann entanglement entropy computed using ED for $N = 12$ sites (thin, light) for various values of Δ shown on a semi-log scale. The spatial bipartition is taken along the central bond. (Inset) The same data on a linear scale for $\Delta/J = 0,0.01$. Dashed lines show fitted linear and logarithmic curves.

which is an XXZ Hamiltonian describing a spin chain with a binary potential set by $q_j = \pm 1$. This XXZ Hamiltonian with quenched disorder serves as one of the paradigmatic models of many-body localization. Although in the context of MBL this model is usually studied with uniformly sampled disorder [25,26], is has also been studied in the case of binary disorder [44,51,52]. Using this mapping and looking at dynamical correlators, we observe the behavior usually found in MBL phases. However, note that our starting point is a disorder-free Hamiltonian (29).

Let us consider the charge density wave initial state as explained in the noninteracting case, see Fig. 10(a). We find that the density imbalance $\Delta\rho$ saturates at a nonzero value (as in the noninteracting case) indicating the persistent memory due to localization. For small interactions, the asymptotic value is close to the one for the noninteracting case for $\Delta = 0.1J$, but as the interaction strength is increased it also acts to stabilize the charge density wave which leads to an increase of the asymptotic value, as can already be seen for $\Delta = 0.3J$. We also observe that the interactions have the effect of damping the fluctuations around this asymptotic value, which is evident even on the short time scales shown in Fig. 10(a).

Next, we can consider the von Neumann entanglement entropy, shown in Fig. 10(b). Here we see a qualitative change (compared to the noninteracting case) in the entanglement growth following the initial area-law plateau. While in the noninteracting case we get linear growth followed by saturation, in presence of density-density interactions, we observe a slower logarithmic growth, as shown by the dashed lines in Fig. 10(b), and in the inset. This logarithmic behavior, which sets in at times $\sim\Delta/J$, is consistent with the phenomenology of MBL.

Let us now consider another term which can be added to the Hamiltonian, and which commutes with the charges, namely the longitudinal field. The latter no longer commutes with plaquette operators and thus the duality mapping (5) is no longer useful. However, we still have the conserved quantities $\hat{q}_j = \hat{\sigma}^x_{j-1,j}\hat{\sigma}^x_{j,j+1}(-1)^{\hat{n}_j}$. In the original spin picture, the effect of this term is to confine spin excitations [53].

In Fig. 11(a), we present the results for entanglement entropy with the Hamiltonian having a longitudinal field term whose strength is controlled by B_x. The results show a rich behavior. In particular, one can notice two new qualitative features. For small $B_x/J \sim 0.2$, we observe logarithmic entanglement growth at a time scale set by B_x/J similar to the MBL behavior observed above. However, for larger $B_x/J \sim 0.5$, we find a slower growth, which can be fit with $\ln(\ln(t))$ as shown in the inset. Furthermore, for small h, the interactions generate extra entanglement compared to the noninteracting results, whereas for larger B_x the entanglement is reduced. We also see this behavior in the density imbalance starting from a charge density wave, shown in Fig. 11(b). We again see that interactions have a damping effect on the fluctuations and that a strong enough field stabilizes the charge density wave.

B. Dynamical charges and quasi-MBL

In the second category of terms, i.e., those terms that do not commute with charges and generate their dynamics, we consider three types of terms:

$$h_z \sum_j \hat{\sigma}^z_{j-1,j}\hat{\sigma}^z_{j,j+1}, \quad B_z \sum_j \hat{\sigma}^z_{j,j+1}, \quad \epsilon \sum_{\langle ij\rangle} \hat{f}_i^\dagger \hat{f}_j. \tag{31}$$

The localization behavior discussed above relies on the presence of static charges $\hat{q}_j$, which act as an effective disorder. It is therefore a natural question to ask what happens when we give dynamics to the charges.

Figure 12 shows the effect of these additional interactions on the density imbalance after a quench from the charge density wave initial state. Figures 12(a) and 12(b) clearly show that the introduction of B_z and ϵ leads to the decay of this imbalance, and ultimately to the disappearance of the charge density wave. One can also see that the time scale at which the results significantly deviate from the $B_z = \epsilon = 0$ case is determined by B_z^{-1} and ϵ^{-1}, respectively. A qualitative difference between these two terms is that the fermion hopping ϵ to the lowest order modifies J. Its effect can be seen in a decrease of the period

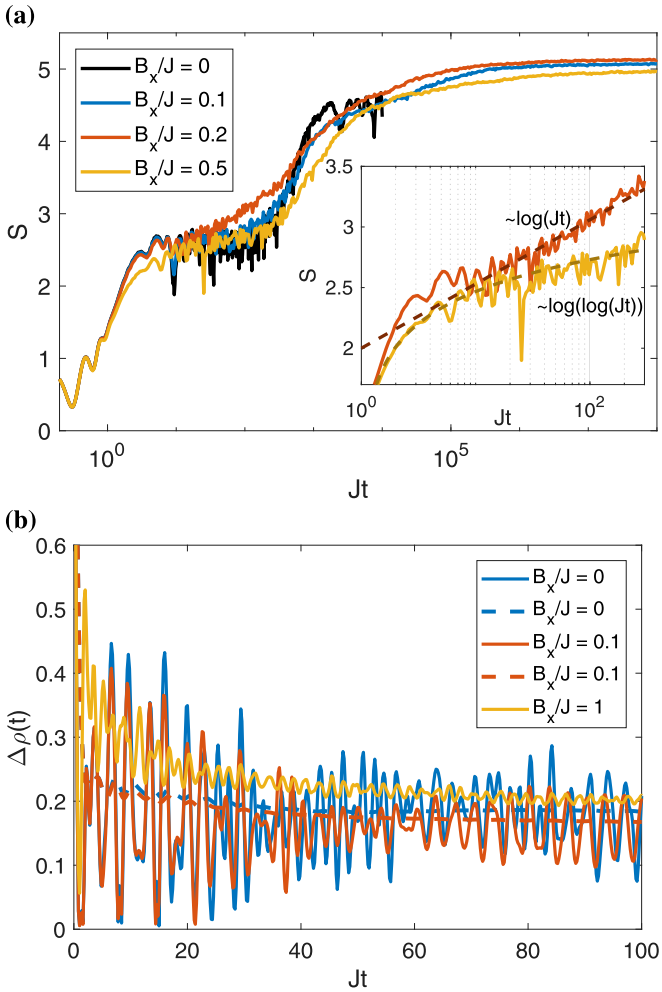

FIG. 11. (a) Entanglement entropy after a quantum quench for various values of the longitudinal field B_x shown on a semi-log scale. A window of the same data for $B_x/J = 0.2,0.5$ is given in the inset. Dashed lines correspond to $\ln(t)$ and $\ln(\ln(t))$ behavior. (b) Density imbalance after a quench from a charge density wave. Results obtained using ED.

of the oscillations at $\epsilon = 0.1$. Beyond this point, there are only quantitative differences between the two cases, and the time averaged values look similar to the eye.

A different phenomenology is observed in the case of the z-Ising coupling h_z, shown in Fig. 12(c). In this case there is little appreciable deviation in the time-averaged value for $h_z/J = 0.001,0.01$, beyond the damping of the oscillations. When the coupling is increased to $h_z = 0.1J$, we finally see appreciable decay of the imbalance, but that does not convincingly vanish, compared to the case for other terms. This can be understood by considering the (anti-)commutation relations of $\hat{\sigma}^z_{j,j+1}$ with the charges $\hat{q}_j$, which are

$$\begin{aligned} \{\hat{\sigma}^z_{j,j+1},\hat{q}_k\} &= 0, \quad k = j, \quad j+1, \\ [\hat{\sigma}^z_{j,j+1},\hat{q}_k] &= 0, \quad k \neq j, \quad j+1. \end{aligned} \tag{32}$$

We can then make an identification with spin operators $\hat{q}_j \to \hat{q}^z_k$ and $\hat{\sigma}^z_{j,j+1} \to \hat{q}^x_j \hat{q}^x_{j+1}$, that is, the transverse field induces an effective Ising coupling between the charges. For the transverse field, we are then precisely in the framework of heavy-light mixtures, which are generally believed to become ergodic at long times. However, the Ising coupling $\hat{\sigma}^z_{j-1,j}\hat{\sigma}^z_{j,j+1}$ maps to the next-nearest neighbor Ising coupling $\hat{q}^x_{j-1}\hat{q}^x_{j+1}$ for the charges. Since the lattice is bipartite, the charges interact separately on these two disconnected sublattices. Importantly, this means that the charges—and by extension the effective disorder potential—on neighboring sites do not become correlated. This leads to increased persistence of localization seen in Fig. 12(c).

Since we still have a heavy-light mixture with the addition of the z-Ising coupling, we can ask whether this additional persistence survives in the thermodynamic limit at long times, which would be in contrast with the standard phenomenology of these systems. Figure 13 shows density imbalance as a function of the system size, which seems to suggest that we also lose localization in this case in the thermodynamic limit, consistent with Ref. [32].

V. DISCUSSION

In this paper, we have studied a family of unconstrained $\mathbb{Z}_2$ lattice gauge theories of spinless fermions coupled to spins-1/2. Using these models, we have revealed a general phenomenology of disorder-free mechanism for localization that extends beyond the models we discussed here. The key feature of this mechanism is the presence of an extensive set of local

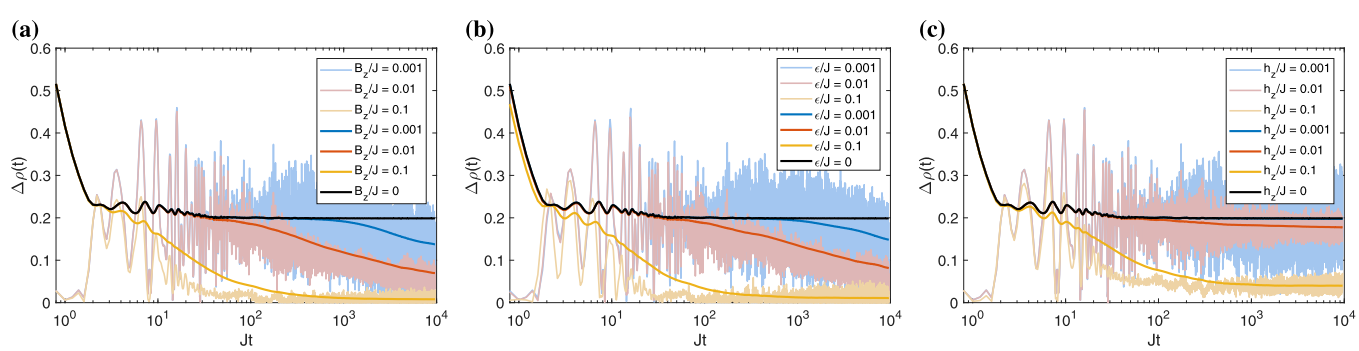

FIG. 12. Density imbalance after a quantum quench from a CDW with dynamical charges. We study a system with $N = 10$ sites with periodic boundary conditions and $h = J$. The light curves are the imbalance $\Delta\rho$ and the dark thick lines are the time averaged value $\frac{1}{t}\int_0^t d\tau\, \Delta\rho(\tau)$. (a) With an additional transverse field $B_z \sum_i \hat{\sigma}^z_{i,i+1}$. (b) Additional fermion hopping $\epsilon \sum_{\langle ij\rangle} \hat{f}^\dagger_i \hat{f}_j$. (c) Added transverse Ising coupling $h_z \sum_i \hat{\sigma}^z_{j-1,j}\hat{\sigma}^z_{j,j+1}$. Dynamics is computed using Krylov subspace methods (Appendix B).

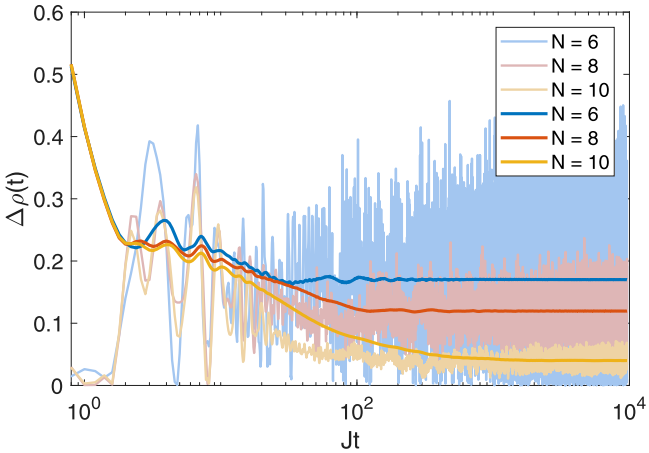

FIG. 13. Finite size scaling of the asymptotic density imbalance in presence of additional Ising coupling $h_z \sum_i \hat{\sigma}^z_{j-1,j}\hat{\sigma}^z_{j,j+1}$. We use system sizes $N = 6,8,10$ with $h_z = 0.1J$ and $h = J$. Light curves show the density imbalance $\Delta\rho$ and the dark and thick curves show the time averaged value of $\frac{1}{t}\int_0^t \mathrm{d}\tau\,\Delta\rho(\tau)$. See Appendix B for details of the Krylov subspace numerical method used.

conserved quantities that emerge from the coupling between the spins and the fermions, with fermions attached to defects in the lattice gauge theory. In particular, we have outlined the connection to the toric code when our model is defined on a periodic square lattice and the resulting constrained dynamics of defects therein. The coupling to fermions allows us to change the behavior of defects depending on the fermion configuration and filling.

As a diagnostic of localization, we have considered the dynamics of the system after a global quantum quench, which revealed the persistence of memory of initial states in both one and two-dimensional case. The results for the density of states and for the localization length confirmed the localization picture. Through these numerical experiments and the exact duality mapping of the Hamiltonian and states, we have demonstrated that a disorder-free mechanism that we found in 1D in Refs. [36,37] applies much more generally.

The effective disorder that appears in our models has a binary nature in contrast to the continuously sampled disorder that is usually studied in the context of localization physics. While for small effective disorder strengths the localization picture for binary disorder and uniform disorder should agree qualitatively, in the strong-disorder limit, we reveal new features that are specific to binary disorder. As most clearly demonstrated in the 1D case, for strong disorder, we find that the single-particle spectrum splits up into a discrete set of degenerate energy levels, of which only very few carry the majority of the spectral weight. This gives rise to resonances which result in the observed area-law plateau in the entanglement entropy, and in long-lived fluctuations of the fermion density. These effects are somewhat washed-out with increased dimensionality. Interestingly, in the 2D case, the strong-disorder limit can be understood as a quantum site percolation problem. By biasing the distribution of the conserved charges above the percolation threshold, we found a delocalization transition of fermions by the percolation mechanism. This behavior was demonstrated by studying a quantum quench from a domain wall initial state, the density of states, and the localization length, which confirmed a direct correspondence with the threshold for classical site percolation. Importantly, our delocalization transition does not require spatial correlations in the disorder, compared to previously known cases. Due to the lower percolation threshold in 3D, for the cubic lattice, we expect that one can observe two delocalized phases for small and large effective disorder h/J, separated by an intermediate localized phase.

While the model we consider in Eq. (1) can be mapped to a model of free fermions, we also considered effects of terms which render the model truly interacting in one dimension. These terms fall into two categories—those that commute with the charges and those that do not. In the first category, we studied nearest-neighbor density interactions, and a longitudinal field. In both cases, in certain parameter regimes, we observed the logarithmic entanglement growth characteristic of many-body localization. Furthermore, for the longitudinal field, we also observed sublogarithmic growth, which shows a good fit with $\ln(\ln(t))$ behavior.

The second class of integrability-breaking terms are those which give dynamics to the conserved charges. These terms generate dynamics similar to that studied in heavy-light mixtures. The latter showed quasi-MBL behavior, i.e., only transient localization. We indeed observe quasi-MBL in the region of parameters which we study. However, we find a special case that was comparatively less effective at destroying the localization behavior. This was due to the dynamics being along the two disconnected chains in the bipartite lattice, which means that nearest-neighbor correlations were not generated. Despite this, it also appears to lead to ergodic behavior in the thermodynamic limit at long times.

The models that we discuss in this paper are particular relevant and timely because of the recent experimental progress in the control of isolated quantum systems in cold atoms experiments [22,23]. Similar to the recent simulation of the Schwinger model [18], our models can be implemented with current technological capabilities. Furthermore, they provide minimal models for lattice gauge theories where in the non-interacting case we are able to perform large scale numerical simulations. Therefore they can serve as a benchmark about which we can consider the truly interacting perturbations, which are practically impossible to simulate in anything other than finite 1D chains.

ACKNOWLEDGMENTS

We are grateful to C. Castelnovo for enlightening discussions. A.S. would like to acknowledge the EPSRC for studentship funding under Grant No. EP/M508007/1. R.M. was in part supported by DFG under Grant SFB 1143. The work of D.L.K. was supported by EPSRC Grant No. EP/M007928/2.

APPENDICES

In the following appendices, we present details of the numerical methods utilized in our calculations. We also include a discussion of the parameters used to produce the figures shown in the main text. In Appendix A, we outline the method for calculating fermionic correlators using determinants for

the case of bilinear Hamiltonians. We explain how to use the Krylov subspace method for time evolution in Appendix B. In Appendix C, we present an outline of the kernel polynomial method. And in Appendix D, we discuss calculations of localization length using transfer matrix techniques.

APPENDIX A: CALCULATION OF FERMION CORRELATORS USING DETERMINANTS

In the case of a Hamiltonian bilinear in fermion operators, dynamic correlation functions can be obtained in terms of determinants. In the main text, we have shown that general correlators for our set of models can be written in terms of purely fermionic correlators. In the following, we explain how the calculation of this correlators can be reduced to determinants, see, e.g., Ref. [54]. A mapping to the free-fermion Hamiltonian dramatically decreases the computational cost compared with ED, which allows us to reach much larger system sizes.

Generically, we are interested in computing expressions of the form

$$\langle\alpha|\exp\{i\sum_{ij}A_{ij}\hat{c}_i^\dagger\hat{c}_j\}|\beta\rangle, \tag{A1}$$

where A is a Hermitian matrix, and $|\alpha\rangle = \hat{c}_{m_N}^\dagger\cdots\hat{c}_{m_1}^\dagger|\text{vac}\rangle$ and $|\beta\rangle = \hat{c}_{n_N}^\dagger\cdots\hat{c}_{n_1}^\dagger|\text{vac}\rangle$ are fermionic Slater determinants. To proceed with the calculation of (A1), we first use the unitarity of the exponential operator $\hat{U}_A \equiv \exp\{i\sum_{ij}A_{ij}\hat{c}_i^\dagger\hat{c}_j\}$ to rewrite (A1) as

$$\begin{aligned}&\langle\text{vac}|\hat{c}_{m_1}\cdots\hat{c}_{m_N}\hat{U}_A\hat{c}_{n_N}^\dagger\hat{U}_A^\dagger\cdots\hat{U}_A\hat{c}_{n_1}^\dagger\hat{U}_A^\dagger\hat{U}_A|\text{vac}\rangle\\ &\quad=\langle\text{vac}\ |\hat{c}_{m_1}\cdots\hat{c}_{m_N}\hat{\tilde{c}}_{n_N}^\dagger\cdots\hat{\tilde{c}}_{n_1}^\dagger|\text{vac}\rangle,\end{aligned} \tag{A2}$$

where $\hat{\tilde{c}}_j^\dagger \equiv \hat{U}_A\hat{c}_j^\dagger\hat{U}_A^\dagger$ and we use $\hat{U}_A|\text{vac}\rangle = |\text{vac}\rangle$. With the help of the Baker-Hausdorff formula, we obtain

$$\hat{\tilde{c}}_i^\dagger = \sum_j \exp\{iA^T\}_{ij}\hat{c}_j^\dagger \equiv \sum_j U_{A,ij}^T\hat{c}_j^\dagger, \tag{A3}$$

distinguishing between the operator $\hat{U}_A$ and the matrix U_A by a hat. Finally, we insert (A3) into (A2), and use the fermionic anticommutation relations to obtain

$$\langle\alpha|\hat{U}_A|\beta\rangle = \det D, \quad D_{jk} = [U_A]_{n_j m_k}, \tag{A4}$$

with $j,k = 1,\ldots,N$. In other words, we select from the matrix U_A those lines and columns that correspond to occupied states in Slater determinants $|\beta\rangle,|\alpha\rangle$.

This derivation allows for further generalizations. For example, in the case of an arbitrary number of unitary operators, using repeatedly the Baker-Hausdorff formula, we obtain

$$\langle\alpha|\hat{U}_A\hat{U}_B\cdots|\beta\rangle = \det D, \quad D_{jk} = [U_AU_B\cdots]_{n_j m_k}. \tag{A5}$$

This equation is suitable for evaluating correlators similar to the one in Eq. (16). In case of the fermion correlators, e.g., (A1), we need to consider expressions of the following form:

$$C_{kl} = \langle\alpha|\hat{c}_k^\dagger\exp\left\{i\sum_{ij}A_{ij}\hat{c}_i^\dagger\hat{c}_j\right\}\hat{c}_l|\beta\rangle. \tag{A6}$$

By commuting $\hat{c}_k^\dagger$ to the left, and $\hat{c}_l$ to the right, we pick up factors $(-1)^{N-p}$ and $(-1)^{N-q}$, where $m_p = k$ and $n_q = l$ and arrive at

$$\begin{aligned}C_{kl} = (-1)^{p+q}&\langle\text{vac}|\hat{c}_{m_1}\cdots\hat{c}_{m_{p-1}}\hat{c}_{m_{p+1}}\cdots\hat{c}_{m_N}\\ &\times\hat{U}_A\hat{c}_{n_N}^\dagger\cdots\hat{c}_{n_{q+1}}^\dagger\hat{c}_{n_{q-1}}^\dagger\cdots\hat{c}_{n_1}^\dagger|\text{vac}\rangle.\end{aligned} \tag{A7}$$

In this case, we need to remove the q-row and the p-column from the matrix D before taking the determinant and then multiply by the corresponding sign. Specifically, we need the $q-p$ cofactor of D where D is given in Eq. (A4). The final expression for the fermion correlator now can be written in a simple form,

$$C_{kl} = D_{lk}^{-1}\det D, \tag{A8}$$

where $D_{jk} = [U_A]_{n_j m_k}$, $j,k = 1,\ldots,N$.

The free-fermion mapping presented in the main text allows one to extract dynamical correlators for system sizes far beyond exact diagonalization. We can estimate the size of the fermionic Hilbert space at half-filling as $N^{-1/2}2^N$ with the spin degrees of freedom adding another factor of 2^N. Instead of diagonalizing exponentially large matrices, the identification of conserved charges allows us to sample uniformly from $\sim 2^N$ determinants of $N\times N$ matrices, corresponding to different charge configurations. Finally, finite-size scaling as well as exact results (up to $N = 20$) show that the required number of samples for a given accuracy scales polynomially with N. Typically, we sample over 10^3–10^4 charge configurations.

APPENDIX B: KRYLOV SUBSPACE DECOMPOSITION

A major bottleneck for computing dynamics using exact diagonalization is the memory requirement for storing many-body wave functions. Although this can be drastically reduced if the model has conserved quantities, memory is still the limiting factor due to the exponential growth of the Hilbert space dimension. A way around this for computing dynamical quantities is to use a smaller set of basis states to perform the time evolution via short time steps. An optimal basis of states for this method can be identified with the Krylov subspace generated by the Hamiltonian. In this appendix, we provide an outline of this method.

Our goal is to efficiently calculate the time evolution of a quantum state $|\Psi(t)\rangle = e^{-i\hat{H}t}|\Psi\rangle$ for which we use Krylov subspace

$$\mathcal{K}_R = \text{span}\{|\Psi\rangle,\hat{H}|\Psi\rangle,\hat{H}^2|\Psi\rangle,\ldots,\hat{H}^{R-1}|\Psi\rangle\}, \tag{B1}$$

where $\hat{H}$ is our Hamiltonian, $|\Psi\rangle$ is an initial state, and R is the number of states in Krylov subspace (that we choose). The idea is that at short enough times the state $|\Psi(t)\rangle$ will be predominantly in this subspace as can be seen from a Taylor expansion of the unitary time evolution

$$|\Psi(t)\rangle = \sum_{n=0}^{\infty}\frac{(-i\hat{H}t)^n}{n!}|\Psi\rangle. \tag{B2}$$

Given a basis for the Krylov subspace $\{|v_i\rangle\}$, the best approximation to the unitary time evolution is given by $e^{-i\tilde{H}t}$ with

$$\tilde{H} = \sum_{ij}|v_i\rangle\langle v_i|\hat{H}|v_j\rangle\langle v_j|. \tag{B3}$$

Therefore we need to diagonalize a matrix of dimension R with matrix elements $\widetilde{H}_{ij} = \langle v_i|\hat{H}|v_j\rangle$. For a Hermitian operator, the reduced Hamiltonian $\widetilde{H}$ takes a simpler tridiagonal form, which can be efficiently diagonalized.

One of the main practical considerations is the accuracy of the Krylov subspace method and the orthogonality between Krylov basis vectors. Due to limited numerical precision, the computed Krylov eigenvectors will diverge from the true ones as more of them are included. A way around this is to orthogonalize each new vector to the previous set. However, the numerical errors that arise in this procedure make it also problematic, and eventually orthogonality will be lost. To get around these issues, after roughly 25 applications of matrix multiplications (the number is chosen empirically) we orthonormalize the entire set of vectors using efficient QR decomposition before proceeding further.

The accuracy of this approximation can be kept below a prescribed threshold only for a finite value of t, which is set by the size of the basis dimension R. To study time evolution on longer timescales, we use a "restarted evolution method," which computes the time evolution up to a certain time δt and then repeats with the new starting state $|\Psi(t+\delta t)\rangle$ until the desired time is reached. To check the accuracy of the method, we then perform the reverse time evolution and check the difference between the values on the forward and backwards pass, which provides a good estimate of the deviation from the true value [55,56].

The method described in this Appendix is limited by the memory required to store the Hamiltonian $\hat{H}$ and the Krylov basis states. Since the Hamiltonian is typically sparse and has $O(\alpha N)$ nonzero values, the memory requirement scales as $O((R+\alpha)N)$ compared with $O(N^2)$ for exact diagonalization. In our calculations, we take $R = 50$, and $\delta t = 1.2$ and compute values for $dt = 0.2$, which we find gives acceptable errors of only 1–2 orders of magnitude above machine precision on the scale of the full time evolution. We note that the computational cost also scales linearly with the number of time steps.

APPENDIX C: DENSITY OF STATES: KERNEL POLYNOMIAL METHOD

The kernel polynomial method (KPM) is a numerical technique for computing spectral quantities. It uses a decomposition into Chebyshev polynomials with modifications to the coefficients to damp Gibbs oscillations which are inherent to Fourier expansion. The benefit of the KPM is that calculations can be reduced to repeated multiplications of the Hamiltonian matrix, which is very efficient for sparse matrices. We will briefly describe the basic procedure, also see Ref. [47] for more details and examples.

1. Chebyshev expansion and modified moments

Chebyshev polynomials are defined on the interval $[-1,1]$ and form the orthogonal basis with respect to inner products defined on this interval with a special weight function. For Chebyshev polynomials of the first kind $T_n(x)$, this weight function is $w(x) = (\pi\sqrt{1-x^2})^{-1}$, so that the inner product reads

$$\langle f|g\rangle_1 = \int_{-1}^{1} \mathrm{d}x\ \frac{f(x)g(x)}{\pi\sqrt{1-x^2}}, \quad \langle T_n|T_m\rangle = \frac{1+\delta_{n,0}}{2}\delta_{n,m}. \tag{C1}$$

Chebyshev polynomials of the second kind $U_n(x)$ are defined with respect to the weight function $w(x) = \pi\sqrt{1-x^2}$, i.e.,

$$\langle f|g\rangle_1 = \int_{-1}^{1} \mathrm{d}x\ \pi\sqrt{1-x^2} f(x)g(x), \quad \langle U_n|U_m\rangle = \frac{\pi^2}{2}\delta_{n,m}. \tag{C2}$$

These polynomials obey useful recursion relations

$$\begin{aligned} T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x), \\ U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x), \end{aligned} \tag{C3}$$

where $T_0(x) = 1$, $T_1(x) = x$, and $U_0(x) = 1$, $U_{-1}(x) = 0$.

Given an orthogonal basis together with the inner product, we can expand a function defined on the interval $[-1,1]$ as

$$f(x) = \alpha_0 + 2\sum_{n=1}^{\infty} \alpha_n T_n(x), \tag{C4}$$

where the moments are given by

$$\alpha_n = \langle f|T_n\rangle_1 = \int_{-1}^{1} \mathrm{d}x\ \frac{f(x)T_n(x)}{\pi\sqrt{1-x^2}}. \tag{C5}$$

However, the numerical integration of the moments is problematic due to the square root appearing in the denominator. To get around this, we instead define the functions

$$\phi_n(x) = \frac{T_n(x)}{\pi\sqrt{1-x^2}}, \tag{C6}$$

which have the property $\langle\phi_n|\phi_m\rangle_2 = \langle T_n|T_m\rangle_1$, and thus we can instead write the expansion as

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}}\left[\mu_0 + 2\sum_{n=1}^{\infty} \mu_n T_n(x)\right], \tag{C7}$$

where the moments are given by

$$\mu_n = \langle f|\phi_n\rangle_2 = \int_{-1}^{1} \mathrm{d}x\ f(x)T_n(x). \tag{C8}$$

To use this expansion in terms of Chebyshev polynomials, we must first rescale the energies and the Hamiltonian so that the bandwidth lies in the interval $[-1,1]$. We thus define the rescaled Hamiltonian and energies

$$\tilde{H} = (H-b)/a, \quad \tilde{E} = (E-b)/a, \tag{C9}$$

where $a = (E_{\max} - E_{\min})/(2-\epsilon)$ and $b = (E_{\max} - E_{\min})/2$. We include a small factor ϵ to avoid stability problems near ± 1. In practice, we can use analytically obtained bounds on $E_{\max/\min}$ to avoid computing them explicitly. One could also compute $E_{\max/\min}$ for smaller system sizes, add a margin of error, and use these for our bounds.

We now have to discretize the function argument and truncate the infinite sum. To make use of the properties of Chebyshev function, we choose a set of K points,

$x_k = \cos(\pi(k+1/2)/K)$, for $k = 0, \ldots, K-1$. With this choice the expansion takes the form

$$f(x_k) = \frac{1}{\pi\sqrt{1-x_k^2}}\left[\mu_0 + 2\sum_{n=1}^{\infty}\mu_n \cos\left(\frac{\pi n(k+1/2)}{K}\right)\right]. \tag{C10}$$

Since we are expanding in periodic functions, the truncation of the sums leads to Gibbs oscillations. If we keep the first M terms in the sum, then to remove these oscillations, we introduce a kernel of order M,

$$K_M(x,y) = g_0\phi_0(x)\phi_0(y) + 2\sum_{m=1}^{M-1} g_m\phi_m(x)\phi_m(y), \tag{C11}$$

which we use to define

$$f_{\mathrm{KPM}}(x) = \int_{-1}^{1} \mathrm{d}y\, \pi\sqrt{1-y^2}K_M(x,y)f(y). \tag{C12}$$

We can then determine the coefficients g_m in the kernel by demanding that f_{KPM} is as close as possible to the true function $f(x)$. Closeness can be defined in a number of different ways each of which lead to different set of coefficients. In our calculations, we use the Jackson kernel defined by coefficients

$$g_m = \frac{1}{M+1}\left[(M-m+1)\cos\left(\frac{\pi m}{M+1}\right) + \sin\left(\frac{\pi m}{M+1}\right)\cot\left(\frac{\pi}{M+1}\right)\right]. \tag{C13}$$

See Ref. [47] for the derivation of these coefficients and discussions of other choices of kernel. Kernel coefficients are then used to modify the moments in our expansion, and we arrive to an expression

$$f(x_k) \approx \frac{1}{\pi\sqrt{1-x_k^2}} \times \left[g_0\mu_0 + 2\sum_{m=1}^{M-1} g_m\mu_m \cos\left(\frac{\pi m(k+1/2)}{K}\right)\right]. \tag{C14}$$

2. Calculation of the moments

The moments that appear in our expansion are typically of the form $\langle\beta|AT_n(H)|\alpha\rangle$, where H is the Hamiltonian matrix, A is a matrix representing an operator and $|\alpha\rangle$ and $|\beta\rangle$ are two states. We need to compute $|\alpha_n\rangle \equiv T_n(H)|\alpha\rangle$, which can be done using recursion relations

$$|\alpha_{n+1}\rangle = 2H|\alpha_n\rangle - |\alpha_{n-1}\rangle, \tag{C15}$$

with $|\alpha_0\rangle = |\alpha\rangle$ and $|\alpha_1\rangle = H|\alpha\rangle$. If $\beta = \alpha$ and $A = \mathrm{I}$, we can use recursion relations for Chebyshev polynomials, specifically, $2T_m(x)T_n(x) = T_{m+n}(x) + T_{m-n}(x)$, to get

$$\mu_{2n} = 2\langle\alpha_n|\alpha_n\rangle - \mu_0, \quad \mu_{2n+1} = \langle\alpha_{n+1}|\alpha_n\rangle - \mu_1, \tag{C16}$$

which reduces the number of matrix operations by approximately half.

We also need to compute moments which involve a trace over states. The latter can be computed efficiently using

$$\mathrm{Tr}[AT_n(H)] \approx \frac{1}{R}\sum_{n=1}^{R-1}\langle r|AT_n(H)|r\rangle, \tag{C17}$$

where D is the size of the matrix H, and $R \ll N$ is the number of chosen random vectors. In all of our calculations we take $R = 30$. The random vectors $|r\rangle$ are defined through random variables ε_{ri},

$$|r\rangle = \sum_{i=0}^{N-1}\varepsilon_{ri}|i\rangle, \tag{C18}$$

where $|i\rangle$ are the basis vectors with an identity in the ith entry. The random variables must satisfy the relation

$$\langle\langle\varepsilon_{ri}\rangle\rangle = 0, \quad \langle\langle\varepsilon_{ri}\varepsilon_{r'j}\rangle\rangle = \delta_{rr'}\delta_{ij}, \tag{C19}$$

that is, they are uncorrelated with zero mean, and have unit mean for their absolute value.

3. Density of states

As an example of the application of the method, let us consider the calculation of the density of states, which is defined as

$$g(E) = \frac{1}{N}\sum_{k=0}^{N-1}\delta(E-E_k). \tag{C20}$$

The coefficients of the Chebyshev expansion are then given by

$$\begin{aligned}\mu_n &= \int_{-1}^{1} dE\ \rho(E)T_n(E) = \frac{1}{N}\sum_{k=1}^{N-1}T_n(E_k)\\ &= \frac{1}{N}\sum_{k=1}^{N-1}\langle k|T_n(H)|k\rangle\\ &= \frac{1}{N}\mathrm{Tr}[T_n(H)],\end{aligned} \tag{C21}$$

which we can compute using the statistical trace and the expectation values as explained above.

In the main text, we use the following parameters for the figures: Fig. 5(a) $\leftarrow N = 10^6$, $M = 1500$, $K = 2M$, $R = 30$; Fig. 5(b) $\leftarrow N = 10^6$, $M = 7500$, $K = 3M$, $R = 30$; Fig. 7(a) $\leftarrow N = (10^3)^2$, $M = 1500$, $K = 2M$, $R = 30$; Fig. 9(b) $\leftarrow N = (10^3)^2$, $M = 2500$, $K = 2M$, $R = 30$; where N is the number of sites, M is the number of moments included in the expansion, K is the number of discretization points, and R is the number of random states used in the statistical trace.

APPENDIX D: LOCALIZATION LENGTH: TRANSFER MATRIX

The application of transfer matrix approach in the calculations of the localization length proceeds by considering a system which is cut up into slices, with slices connected via transfer matrices [48]. From these, we can extract the eigenvalues of a limiting matrix that gives the Lyapunov

exponents for our system. For instance, consider a 1D chain with the Hamiltonian

$$\hat{H} = -J\sum_j(\hat{c}_j^\dagger\hat{c}_{j+1} + \text{H.c.}) - \sum_j h_i\hat{c}_j^\dagger\hat{c}_j. \tag{D1}$$

The action of this Hamiltonian on an eigenstate $|\psi\rangle = \sum_i \psi_i|i\rangle$, where $|i\rangle$ is the state localized on site i gives the relation

$$E\psi_i = -J\psi_{i+1} - J\psi_{i-1} - h_i\psi_i, \tag{D2}$$

where E is the eigenvalue of the state $|\psi\rangle$. This equation can be written in a compact form by introducing a transfer matrix

$$\begin{pmatrix}\psi_{i+1}\\ \psi_i\end{pmatrix} = \begin{pmatrix}-\frac{1}{J}(E+h_i) & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\psi_i\\ \psi_{i-1}\end{pmatrix}. \tag{D3}$$

In higher dimensions, these equations have to be modified slightly, in particular, we get

$$E\psi_i = -J\psi_{i+1} - J\psi_{i-1} - H_{\text{perp}}\psi_i, \tag{D4}$$

where ψ_i is now a vector of the values of the wave function in the orthogonal direction at the slice i, and H_{perp} is the matrix representation of the Hamiltonian in the same direction. The transfer matrix equation assumes the form

$$\begin{pmatrix}\psi_{i+1}\\ \psi_i\end{pmatrix} = \begin{pmatrix}-\frac{1}{J}(E+H_{\text{perp}}) & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\psi_i\\ \psi_{i-1}\end{pmatrix}. \tag{D5}$$

Given the transfer matrix, we can compute the product along a long chain of length L,

$$Q_L = \prod_{i=1}^{L} T_i, \tag{D6}$$

where T_i is the transfer matrix for slice i. Oseledec's theorem states that there exists a limiting matrix

$$\Gamma = \lim_{L\to\infty}(Q_L Q_L^\dagger)^{1/2L}, \tag{D7}$$

with eigenvalues $\exp(\gamma_i)$, where γ_i are the Lyapunov exponents of the matrix Q_L. The smallest Lyapunov exponent describes the slowest growth of the wave function and corresponds to the inverse of the localization length λ.

More intuitively, we can consider Q_L as the transfer matrix between the extreme ends of the chain:

$$\begin{pmatrix}\psi_{L+1}\\ \psi_L\end{pmatrix} = Q_L\begin{pmatrix}\psi_2\\ \psi_1\end{pmatrix}, \tag{D8}$$

so the eigenvalues of Q_L describe the growth of the wave function along the length of the system. We can then take the smallest eigenvalue, q, of the matrix Q_L, and compute the localization length via

$$\lambda = \frac{L}{\ln(q)}. \tag{D9}$$

The procedure of computing a matrix product of a large number of matrices is numerically unstable since the matrix elements diverge or vanish exponentially. We therefore must orthonormalize the product of matrices after a few number steps. The numerical procedure proceeds as follows: we iteratively construct the product matrix Q by applying randomly generated transfer matrices T. After the number of applications exceeds a predefined limit or the amplitude of the elements of the matrix exceed a threshold, we store the logarithm of the eigenvalues and orthonormalize the matrix Q, which can be done efficiently using QR decomposition. We continue applying T, storing eigenvalues and orthonormalizing until we have reached the length of the chain L. See below a pseudocode of the algorithm for computing the localization length λ:

```
Q ← Id
l ← 1
While l ⩽ L do
   count ← 1
   While max(abs(Q)) ⩽ theshold or count ⩽ limit do
      Initialize random T for slice
      Q ← TQ
      l++
      count++
   end While
   b ← eig(Q)
   c ← c + ln(b)
   Q ← orthonormal(Q)
end While
b ← eig(Q)
c ← c + ln(b)
lambda ← max(L/c).
```


[1] J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979).
[2] E. Fradkin, *Field Theories of Condensed Matter Physics* (Cambridge University Press, Cambridge, 2013).
[3] X.-G. Wen, *Quantum Field Theory of Many-Body Systems* (Oxford University Press, Oxford, UK, 2007).
[4] U.-J. Wiese, Ann. Phys. **525**, 777 (2013).
[5] C. L. Henley, Annu. Rev. Condens. Matter Phys. **1**, 179 (2010).
[6] A. Kitaev, Ann. Phys. (NY). **303**, 2 (2003).
[7] A. Kitaev, Ann. Phys. (NY). **321**, 2 (2006).
[8] R. Moessner and S. L. Sondhi, Phys. Rev. Lett. **86**, 1881 (2001).
[9] R. Moessner, S. L. Sondhi, and E. Fradkin, Phys. Rev. B **65**, 024504 (2001).
[10] A. Rüegg, S. D. Huber, and M. Sigrist, Phys. Rev. B **81**, 155118 (2010).
[11] R. Žitko and M. Fabrizio, Phys. Rev. B **91**, 245130 (2015).
[12] R. Nandkishore, M. A. Metlitski, and T. Senthil, Phys. Rev. B **86**, 045128 (2012).
[13] S. A. Parameswaran and S. Gopalakrishnan, Phys. Rev. Lett. **119**, 146601 (2017).
[14] A. E. Antipov, Y. Javanmard, P. Ribeiro, and S. Kirchner, Phys. Rev. Lett. **117**, 146601 (2016).
[15] A. J. Herrmann, A. E. Antipov, and P. Werner, Phys. Rev. B **97**, 165107 (2018)
[16] S. Gazit, M. Randeria, and A. Vishwanath, Nat. Phys. **13**, 484 (2017).

[17] G. Schubert and H. Fehske, Phys. Rev. B **77**, 245130 (2008).
[18] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, Nature (London) **534**, 516 (2016).
[19] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[20] R. Nandkishore and D. A. Huse, Annu. Rev. Condens. Matter Phys. **6**, 15 (2015).
[21] D. A. Abanin and Z. Papić, Ann. Phys. **529**, 1700169 (2017).
[22] J.-y. Choi, S. Hild, J. Zeiher, P. Schauss, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, Science **352**, 1547 (2016).
[23] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Luschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, Science **349**, 842 (2015).
[24] D. Basko, I. Aleiner, and B. Altshuler, Ann. Phys. (NY). **321**, 1126 (2006).
[25] M. Žnidarič, T. Prosen, and P. Prelovšek, Phys. Rev. B **77**, 064426 (2008).
[26] J. H. Bardarson, F. Pollmann, and J. E. Moore, Phys. Rev. Lett. **109**, 017202 (2012).
[27] M. Srednicki, Phys. Rev. E **50**, 888 (1994).
[28] Y. Kagan and L. A. Maksimov, Zh. Eksp. Teor. Fiz. **87**, 348 (1984) [Sov. Phys. JETP **60**, 201 (1984)].
[29] M. Schiulaz and M. Müller, in *15th International Conference on Transport in Interacting Disordered Systems (TIDS15)*, edited by M. Palassini, AIP Conf. Proc. No. 1610 (AIP, New York, 2014), p. 11.
[30] M. Schiulaz, A. Silva, and M. Müller, Phys. Rev. B **91**, 184202 (2015).
[31] N. Y. Yao, C. R. Laumann, J. I. Cirac, M. D. Lukin, and J. E. Moore, Phys. Rev. Lett. **117**, 240601 (2016).
[32] Z. Papić, E. M. Stoudenmire, and D. A. Abanin, Ann. Phys. (NY). **362**, 714 (2015).
[33] M. van Horssen, E. Levi, and J. P. Garrahan, Phys. Rev. B **92**, 100305 (2015).
[34] J. M. Hickey, S. Genway, and J. P. Garrahan, J. Stat. Mech. (2016) 054047.
[35] J. P. Garrahan, Physica A **504**, 130 (2018).
[36] A. Smith, J. Knolle, D. L. Kovrizhin, and R. Moessner, Phys. Rev. Lett. **118**, 266601 (2017).
[37] A. Smith, J. Knolle, R. Moessner, and D. L. Kovrizhin, Phys. Rev. Lett. **119**, 176601 (2017).
[38] J. Janarek, D. Delande, and J. Zakrzewski, Phys. Rev. B **97**, 155133 (2018).
[39] A. Alvermann and H. Fehske, Eur. Phys. J. B **48**, 295 (2005).
[40] G. Schubert and H. Fehske, arXiv:0905.2537.
[41] C. Prosko, S.-P. Lee, and J. Maciejko, Phys. Rev. B **96**, 205104 (2017).
[42] M. Brenes, M. Dalmonte, M. Heyl, and A. Scardicchio, Phys. Rev. Lett. **120**, 030601 (2018).
[43] B. Paredes, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **95**, 140501 (2005).
[44] T. Enss, F. Andraschko, and J. Sirker, Phys. Rev. B **95**, 045121 (2017).
[45] G. Baskaran, S. Mandal, and R. Shankar, Phys. Rev. Lett. **98**, 247201 (2007).
[46] J. Knolle, D. L. Kovrizhin, J. T. Chalker, and R. Moessner, Phys. Rev. Lett. **112**, 207203 (2014).
[47] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, Rev. Mod. Phys. **78**, 275 (2006).
[48] B. Kramer and A. MacKinnon, Reports Prog. Phys. **56**, 1469 (1993).
[49] J. Hauschild, F. Heidrich-Meisner, and F. Pollmann, Phys. Rev. B **94**, 161109 (2016).
[50] D. J. Thouless, J. Phys. C Solid State Phys. **5**, 77 (1972).
[51] F. Andraschko, T. Enss, and J. Sirker, Phys. Rev. Lett. **113**, 217201 (2014).
[52] B. Tang, D. Iyer, and M. Rigol, Phys. Rev. B **91**, 161109 (2015).
[53] M. Kormos, M. Collura, G. Takács, and P. Calabrese, Nat. Phys. **13**, 246 (2016).
[54] D. L. Kovrizhin and J. T. Chalker, Phys. Rev. B **81**, 155318 (2010).
[55] R. B. Sidje, ACM Trans. Math. Softw **24**, 130 (1998).
[56] M. Brenes, V. K. Varma, A. Scardicchio, and I. Girotto, arXiv:1704.02770.