Local integrals of motion detection of localization-protected topological order

Thorsten B. Wahl,1 Florian Venm,1 and Benjamin Béri1,2

1 DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK
2 T.C.M. Group, Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge, CB3 0HE, UK

(Dated: November 2021)

Many-body-localized (MBL) phases can be topologically distinct, but distinguishing these phases using order parameters can be challenging. Here we show how topologically distinct local integrals of motion, variationally parametrized by quantum circuits, can be used to numerically demonstrate the topological inequivalence of MBL phases. We illustrate our approach on a fermionic chain where both topologically distinct MBL phases and benchmark comparisons to order parameters are possible. We also use our approach, augmented by the DMRG-X algorithm, to extract high-energy topological doublets. We describe applying our methods to higher dimensions to identify MBL topological order and topological multiplets hidden by the dense many-body spectrum.

I. INTRODUCTION

Many-body localization (MBL) has attracted a wealth of interest in the last fifteen years. One of the most striking features of MBL is the violation of the eigenstate thermalization hypothesis: MBL systems do not thermalize, but instead retain some memory of their initial state. As a result, some MBL systems are able to protect quantum information.

MBL systems are characterized by local integrals of motion (LIOMs) that are typically assumed to be related to Pauli-$z$ operators $\sigma_i^z$ acting on site $i$ via a local unitary transformation $U_i$, $\tau_i^z = U_i \sigma_i^z U_i^\dagger$. The $\tau_i^z$ are thus exponentially localized. They define a complete set of quantum numbers since, $[H, \tau_i^z] = [\tau_i^z, \tau_j^z] = 0$ for all $i, j = 1, \ldots, N$, with $N$ the system size. Hence, for topological MBL systems the notion of LIOMs has to be extended: One must use tLIOMs, given by $\tau_i = U S_i U^\dagger$, where $U$ is again a unitary, but $\{S_i\}$ is now a set of mutually commuting local stabilizers whose common eigenstates all display (the same) topological order. On a topologically non-trivial manifold, to get a complete set of quantum numbers one augments tLIOMs by the non-local $\tau_i^{nl} = U S_i^{nl} U^\dagger$, where $S_i^{nl}$ are non-contractible Wilson loops (i.e., logical operators).

In this work, we show how a topological LIOM framework, with LIOMs and tLIOMs for topologically trivial and nontrivial cases, respectively, combined with quantum circuits for MBL, can be used to numerically detect topological MBL. Furthermore, as we also show, when used in conjunction with the excited-state density-matrix renormalization group (DMRG-X) this approach can also identify topological multiplets provided the system is deep in a topological MBL phase.

Due to its polynomial scaling with system size, and the generality of tLIOMs, our approach is a general way to capture topological MBL, including beyond 1D. Nonetheless, to demonstrate its use, we focus on 1D: we study the disordered interacting Kitaev chain, displaying two topologically distinct MBL phases. Studying this system is useful not only due to the exact diagonalization (ED) benchmark available in 1D, but also due to a local order parameter benchmark available thanks to a duality to a system displaying conventional symmetry breaking.

II. (TOPOLOGICAL) LIOMS FROM QUANTUM CIRCUITS

LIOMs are typically assumed to be related to Pauli-$z$ operators $\sigma_i^z$ acting on site $i$ via a local unitary transformation $U$, $\tau_i^z = U \sigma_i^z U^\dagger$. The $\tau_i^z$ are thus exponentially localized. They define a complete set of quantum numbers since, $[H, \tau_i^z] = [\tau_i^z, \tau_j^z] = 0$ for all $i, j = 1, \ldots, N$, with $N$ the system size. Hence, for topological MBL systems the notion of LIOMs has to be extended: One must use tLIOMs, given by $\tau_i = U S_i U^\dagger$, where $U$ is again a unitary, but $\{S_i\}$ is now a set of mutually commuting local stabilizers whose common eigenstates all display (the same) topological order. On a topologically non-trivial manifold, to get a complete set of quantum numbers one augments tLIOMs by the non-local $\tau_i^{nl} = U S_i^{nl} U^\dagger$, where $S_i^{nl}$ are non-contractible Wilson loops (i.e., logical operators).

To utilize tLIOMs numerically, we use that owing to its locality, we can efficiently approximate $U$ by a fixed-depth quantum circuit $U_{SR}$. We employ $U_{SR}$ variationally: we aim for $[H, \tau_i^z] = 0$ by minimizing

$$f = \frac{1}{2} \sum_{i=1}^{N} \text{tr} \left( [\tau_i^z, H] [\tau_i^z, H]^\dagger \right),$$

where $\tau_i^z = U_{SR} S_i U_{SR}^\dagger$ and $S_i$ are topological stabilizers for a tLIOM ansatz, while $S_i = \sigma_i^z$ for a conventional LIOM ansatz. The lower $f$, the better the approximate (t)LIOs describe the system. In the topological case, the non-local $\tau_i^{nl}$ enter only for multiplet splittings. Hence, they almost commute with $H$ if all tLIOMs are optimized and thus can be omitted in Eq. 1. To minimize $f$, we proceed similarly to Ref. We expand...
\(H\) as a sum of local terms and identify which combination of local terms contributes to \(f\) (which is quadratic in \(H\)). Each of the contributions is then represented as a tensor network contraction, which we evaluate efficiently using common numerical methods \[34, 42, 43\]. While developed for non-topological systems, this approach directly applies also to the topological case: by using topological \(S_i\) (which act on multiple sites) we merely increase the number of tensors in each contraction. In the non-topological phase, we expect the conventional LIOM ansatz to perform better. In the topological phase, however, the tLIOM ansatz will minimize \(f\).

### III. MODEL

The disordered interacting Kitaev chain \(37\) is a system of Majorana fermions \(\gamma_n = \gamma_n^\dagger\) with \(\{\gamma_m, \gamma_n\} = 2\delta_{mn}\); the Hamiltonian is (cf. Fig. 1a)

\[
H = \sum_{j=1}^{N-1} it_j \gamma_{2j-1} \gamma_{2j} + \sum_{j=1}^{N-1} i\mu_j \gamma_{2j-1} \gamma_{2j+1} + \sum_{j=1}^{N-1} V_j \gamma_{2j-1} \gamma_{2j+1} + \gamma_{2j+2},
\]

(2)

where the tunnel amplitude \(t_j\), the on-site potential \(\mu_j\) and the interaction strength \(V_j\) are Gaussian distributed random variables with zero mean and standard deviation \(\sigma_t = 1, \sigma_\mu\) and \(\sigma_V\), respectively. The Hamiltonian commutes with the fermion parity operator \(Z = \prod_{n=1}^{2N} \gamma_n\), which splits the Hilbert space into two parity sectors.

For \(\sigma_\mu, \sigma_V \gg 1\), the system is in the trivial phase with commuting projector representative

\[
H^{\text{triv}} = \sum_{j=1}^{N} it_j \gamma_{2j-1} \gamma_{2j} + \sum_{j=1}^{N-1} V_j \gamma_{2j-1} \gamma_{2j+1} + \gamma_{2j+2}.
\]

(3)

\(H^{\text{triv}}\) may be expressed in terms of the stabilizers \(S_j^{\text{triv}} = \gamma_{2j-1} \gamma_{2j}, j = 1, \ldots, N\); consequently, its eigenstates correspond to the occupation of fermionic modes associated to physical sites, cf. Fig. 1b.

For \(\sigma_\mu = \sigma_V = 0\) the system is in the topological phase.

\[
H^{\text{topo}} = \sum_{j=1}^{N-1} it_j \gamma_{2j} \gamma_{2j+1},
\]

(4)

with \(S_j^{\text{topo}} = i\gamma_{2j} \gamma_{2j+1}, j = 1, \ldots, N-1\), gives the commuting projector representative of this phase. The eigenstates again correspond to the occupation of fermionic modes, but now they come from pairs of Majorana modes straddling physical sites, cf. Fig. 1c. \(S_n^{\text{triv}} = i\gamma_{2N} \gamma_1\) forms a zero-energy fermion mode, resulting in a two-fold degenerate energy spectrum.

### IV. ED AND ORDER PARAMETER BENCHMARKS

To provide a benchmark for our subsequent tLIOM analysis, we use ED to locate the MBL phases. We first analyze the level spacing for \(N = 14\). In Fig. 2a, we show the gap ratio \(r_n = \min(s_{n-1}, s_n)/\max(s_{n-1}, s_n)\) with \(s_n = E_n - E_{n-1}\) the level spacing in a given parity sector. For each data point we average over the mid-third of energies (as those best reflect whether the system is thermalizing \[27, 49, 50\]) and 100 disorder realizations. In an MBL phase we expect \(r_P = 0.386\) (Poisson distribution) due to the lack of level repulsion; in a thermal phase we expect \(r_{WD} = 0.530\) (Wigner-Dyson distribution). Fig. 2a indicates that there are two MBL phases (with \(r \approx r_P\)), one for \(\sigma_\mu + \sigma_V \lesssim 1\) and another one for \(\sigma_\mu \gtrsim 1\) or \(\sigma_V \gtrsim 1\). Although the gap ratio never gets as large as \(r_{WD}\), between these regions our results are consistent with an extended delocalized phase \[19, 51, 52\].

Topological properties can be detected by the spin-glass order parameter \[10, 27, 53\]

\[
\chi_n^{SG} = \frac{1}{N} \sum_{j=1}^{N} \langle n | \sigma_j^x \sigma_j^y | n \rangle^2
\]

in eigenstates \(|n\rangle\) of the quantum Ising chain linked to our system via Jordan-Wigner transformation (cf. Appendix A). While this is a two-point
correlator of local operators in the spin language, it is a non-local order parameter (with $\gamma_j$ strings), as befits one detecting topological features, for the fermionic system. We also use a “dual order parameter” $\chi_n^{SG} = 1/N \sum_{i,j=1}^{N} \langle n| \tilde{\sigma}_i^z \tilde{\sigma}_j^z |n \rangle^2$, where $\tilde{\sigma}_j^z$ are analogous to disorder operators $\tilde{\sigma}_j^z$ in the Ising chain (cf. Appendix A). $\chi_n^{SG} \propto N$ for $N \gg 1$ in the topological phase, while $\chi_n^{SG} \rightarrow 1$ in the trivial phase. Conversely, $\chi_n^{dSG} \propto N$ for $N \gg 1$ in the trivial phase, while $\chi_n^{dSG} \rightarrow 1$ in the topological phase. Our results for $\chi_n^{dSG}$ (see Fig. 2, c and Appendix B) suggest that the $\sigma_+ + \sigma_- \lesssim 1$ phase is topological, while the other MBL phase is trivial.

V. PHASE DIAGRAM FROM TLIOMS

We now apply the tLIOM approach to the model. We optimize, using the algorithm of Ref. 34, the quantum circuits $U_{SR}$ and $U_{SR}^{(t)}$ (for approximate LIOMs and tLIOMs, respectively) over the space of fermion parity conserving unitaries. We use two-layer quantum circuits with gates acting on $\ell$ sites each. In Fig. 3, we show the resulting normalized figure of merit $f/(\tau(\mathcal{H}^2))$ for conventional (topological) LIOMs, for system size $N = 48$, gate lengths $\ell = 2, 4, 6$ and focusing on $\sigma_+ \doteq \sigma_-$. We also show $f/(\tau)$ for $\ell = 6$ in the entire two-parameter phase diagram. As a function of $\ell$, we see a roughly exponential improvement deep in the MBL phases for the corresponding set of LIOMs.

Another way to map out the phase diagram using (t)LIOMs is entanglement entropy fluctuations $\sigma_{S}$. We cut the approximate matrix product eigenstates, obtained from the optimized (t)LIOM ansätze, at a point at least $3\ell/2$ sites away from the boundary and apply the algorithm of Ref. 34 to compute the average entanglement entropy. (Here, we averaged over all approximate eigenstates, as their entropies depend only on the expectation values of the tLIOMs near the cut [34], which are not directly linked to overall energies.) We repeat this step for all disorder realizations and calculate the standard deviation $\sigma_S$ of these averages. Finally, to reduce statis-
tical fluctuations, we average $\sigma_S$ over all cut positions to obtain $\langle \sigma_S \rangle_{\text{cuts}}$.

We expect $\langle \sigma_S \rangle_{\text{cuts}} \to 0$ for $\sigma_V = \sigma_\mu \to \infty$, because there the LIOMs are simply the trivial stabilizers (exact LIOM ansatz with $U_{\text{SR}} = 1$): the eigenstates are independently populated local fermion modes. Similarly, $\sigma_\nu = \sigma_u = 0$ is the topological MBL limit with tLIOMs simply the topological stabilizers (exact tLIOM ansatz with $U_{\text{SR}}^{(t)} = 1$), independently of the disorder realizations, which again implies $\langle \sigma_S \rangle_{\text{cuts}} = 0$. Away from these limits, but still in an MBL phase, i.e., with all eigenstates MBL, the entanglement and its fluctuation are low due to the area law. In an ergodic phase, where all eigenstates are volume-law entangled, the equivalence of spectral and ensemble (i.e., over disorder) averages [55] implies suppressed entanglement fluctuations. In a thermal but not ergodic phase, with a $0 < \nu < 1$ fraction of volume-law states, we expect entanglement fluctuations to diverge in the thermodynamic limit due to $\nu$ being sensitive to small changes in the disorder [27].

The maximum amount of entanglement allowed by our approximation is proportional to $\ell$; hence, we expect $\langle \sigma_S \rangle_{\text{cuts}}$ to acquire a maximum, increasing with $\ell$ [34], where entanglement fluctuations would diverge. (This is analogous to the finite-size scaling in Ref. [27].) This is consistent with Fig. 3 if one combines the LIOM and tLIOM data in the regimes where they are reliable.

Based on the comparison between the optimized figures of merit for $\sigma_\mu = \sigma_V$, we expect the topological MBL phase to extend at most up to $\sigma_\nu \lesssim 0.5$ and the trivial MBL phase to start above $\sigma_\mu = \sigma_V \gtrsim 0.5$. But the broad maxima of the entanglement entropy fluctuation of $U_{\text{SR}}^{(t)}$ at $\sigma_\mu = \sigma_V \approx 0.5$ can indicate an interstitial thermal phase. This is also consistent with conventional and topological LIOMs performing comparably in a window around $\sigma_\nu = \sigma_V \approx 0.5$. Outside of this window, however, the performance difference between conventional and topological LIOM ansätze allows us to clearly detect the topology of the MBL phase.

The picture obtained from our (t)LIOM approach is thus consistent with the (dual) spin-glass order parameter and the gap ratio from ED. While meeting this benchmark, using (t)LIOMs we could probe system sizes well beyond the reach of ED. This suggests that our method can generalize well for detecting topological MBL phases beyond 1D.

---

VI. TOPOLOGICAL DOUBLETS

Another topological MBL feature is the presence of topological multiplets across the entire energy spectrum [15, 16]. (The multiplet splitting is exponentially small in system size.) In the Kitaev chain, one has topological doublets [16, 37]. Here we show how these can be detected using tLIOMs. A key ingredient is the access to the approximate nonlocal $\tau^{nl} = U_{\text{SR}}^{(t)}S^{nl}U_{\text{SR}}^{(t)\dagger}$ and the corresponding flip operator (an MBL strong zero mode [56]) $\tau^{nl,x} = U_{\text{SR}}^{(t)}\gamma_1U_{\text{SR}}^{(t)\dagger}$: the approximate eigenstate $|\psi_\alpha\rangle$ forms a doublet with $\tau^{nl,x}|\psi_\alpha\rangle$. The goal is to show that these two states have approximately the same energy.

To enhance energy accuracy, and to get a method that can access spatial features beyond $\ell$ (which is required for studying multiplet splittings in large systems, see Appendix C), we couple the tLIOM ansatz with DMRG-X [35]: we write $|\psi_\alpha\rangle$ and $\tau^{nl,x}|\psi_\alpha\rangle$ as matrix product states to use as inputs to DMRG-X. We denote the resulting outputs by $|\Psi_\alpha\rangle$ and $|\Psi_\alpha^l\rangle$, respectively. As DMRG-X is biased towards low-entanglement states [57], we must prevent convergence to $|\Psi_\alpha^l\rangle$ with different tLIOM expectation values than $|\psi_\alpha\rangle$, but with lower entanglement.

To control for this, a convenient proxy is to require high $\sqrt{\langle |\Psi_\alpha\rangle|\langle \Psi_\alpha^l|\tau^{nl,x}|\Psi_\alpha\rangle\rangle}$ which for tensor network states are broadly interpretable as geometric mean “per-site overlaps.”

We now turn to demonstrating this approach. To be able to compare to ED, we work with $N = 15$ and deep in the topological MBL phase so that ED can assign almost all doublets correctly due to their small level splitting. We optimize tLIOMs for 10 disorder realizations and for each realization we randomly choose 10 approximate eigenstates and their doublet partners. In DMRG-X, we use bond dimensions up to 85. Our DMRG-X-augmented tLIOM results, direct tLIOM and ED data are shown in Fig. 4. The DMRG-X-augmented tLIOM splittings are in quantitative agreement with ED up to $\sigma_\nu = \sigma_\mu \approx 0.13$. Beyond this, the breakdown of the method is signaled by the deteriorating per-site overlaps.

While we demonstrated its use for system sizes where a comparison to ED is available, the per-site overlap is expected to be a useful and feasibly computable indicator for system sizes beyond the reach of ED.)

Deep in the topological MBL phase, we expect this method to correctly predict the splittings for $N$ beyond the reach of ED. The required bond dimensions will be larger, but still scale polynomially with $N$ due to having area-law eigenstates. Beyond 1D, one could similarly take an approximate eigenstate obtained from the tLIOM approach [58, 59] in the topological phase as a starting point for an approximation using projected entangled pair states (PEPS) [60] methods [61]. Although the mean level spacing is exponentially smaller than the multiplet splitting [33], the locality of PEPS will allow
We showed how tLIOMs can detect topological MBL in numerical simulations. Using (t)LIOM numerics for the interacting disordered Kitaev chain (up to $N = 48$), we found two topologically distinct MBL phases and features consistent with an intervening thermal phase. We also showed how to detect topological multiplets using tLIOMs which, for strong MBL and when augmented by DMRG-X, reached very high energy accuracy. It would be interesting to use tLIOMs above 1D (e.g., for the toric code) to detect MBL topological order and to detect topological multiplets, which here have splitting decaying slower with system size than the mean level spacing even for strong MBL.

Acknowledgments.—We thank Nicolas Laflorencie for very helpful discussions. This work was supported by the EPSRC grant EP/S019324/1 and the ERC Starting Grant No.678795 TopInSy. Our computations used resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the EPSRC (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).

Appendix A: Ising chains via Jordan-Wigner transformations

We rewrite the Hamiltonian $H$ using a Jordan-Wigner transformation, $\gamma_{2j} = (\prod_{k=1}^{j} \sigma_k^x) \sigma_j^y$, $\gamma_{2j+1} = - (\prod_{k=1}^{j} \sigma_k^x) \sigma_j^y$ (see Ref. [3]). Its authors introduced a real-space renormalization group method for excited states, predicting two distinct MBL phases separated by a transition visible in the entire energy spectrum.

An alternative Jordan-Wigner transformation

$$H = \sum_{j=1}^{N-1} t_j \sigma_j^x \sigma_{j+1}^x + \sum_{j=1}^{N} \mu_j \sigma_j^y - \sum_{j=1}^{N-1} V_j \sigma_j^x \sigma_{j+1}^x$$

with Pauli operator $\sigma_i^\alpha$, $\alpha = x, y, z$ acting on site $i$. Due to fermion parity conservation, the Hamiltonian is invariant under the parity operator $Z = \prod_{i} \sigma_i^z$. This generalized quantum Ising Hamiltonian is the same as the $\alpha \hbar J$ model studied in Ref. [3].

One alternative Jordan-Wigner transformation

$$\gamma_{2j} = (\prod_{k=1}^{j} \sigma_k^y) \sigma_j^x$$

results in Kramers-Wannier-like duality, such that

$$H = \sum_{j=1}^{N} N_{\sqrt{\text{overlap}}} + \sum_{j=2}^{N} \mu_j \sigma_j^z - \sum_{j=2}^{N} V_j \sigma_j^z \sigma_{j+1}^z$$

with $\bar{Z} = \prod_{j=1}^{N} \sigma_j^z = -Z$ and $\sigma_j^y$ likewise Pauli operators, but now acting on links. The two sets of Pauli operators are related via the transformation

$$\sigma_j^x = \begin{cases} \left(\prod_{k=1}^{j} \sigma_k^y \right) \sigma_N^y & \text{if } j < N, \\ \sigma_N^y & \text{if } j = N, \end{cases}$$

$$\sigma_j^z = \begin{cases} \sigma_j^z & \text{if } j > 1, \\ -Z \sigma_1^y \sigma_N^y & \text{if } j = 1, \end{cases}$$

and inverse transformation

$$\bar{\sigma}_j = \begin{cases} -\sigma_j^y \bar{\sigma}_{j+1} & \text{if } j > 1, \\ -\sigma_j^y & \text{if } j = 1, \end{cases}$$

(An actual Kramers-Wannier transformation, where the $-\sigma_j^y$ in Eq. (A5) are replaced by $\sigma_j^x$ and the second line in Eq. (A6) by $\sigma_N^y$, changes only the last two terms of Eq. (A2), which become local boundary terms.) In each symmetry sector of $\bar{Z}$ (even / odd parity) Eq. (A2) is a generalization of the quantum Ising model with next-nearest neighbor coupling considered in Ref. [27] (after an
on-site unitary transformation sending \( \sigma_f^x \leftrightarrow \sigma_f^z \). There, the authors likewise found two distinct MBL phases as evidenced by spin-glass order. We note that for \( \mu_j = 0 \) \( \forall j = 1, \ldots, N \) the Hamiltonian in Eq. (A2) decouples in each parity sector into two non-interacting chains, cf. the structure of Eq. (A1) for \( \sigma_V = 0 \). Hence, there is a phase transition point at \( (\sigma_V, \mu) = (1, 0) \).

The bases \( \{\sigma_\alpha^n\} \) and \( \{\sigma_\gamma^n\} \) are related via a unitary transformation exchanging the standard error of the mean. The bases \( \{\sigma_\alpha^n\} \) and \( \{\sigma_\gamma^n\} \) are related via a unitary transformation exchanging stabilizer sets of complementary topological character (see Fig. 5), while \( \chi_n^{SG} \propto N \) (for \( N \gg 1 \)) for MBL with topological order (and \( \chi_n^{SG} \rightarrow 1 \) otherwise), we have \( \chi_n^{dSG} \propto N \) (for \( N \gg 1 \)) for topologically trivial MBL (and \( \chi_n^{dSG} \rightarrow 1 \) otherwise). This behavior can be seen in Fig. 6 where we plot the dual order parameter as a function of \( \chi = \mu \).

**Appendix C: Quantum circuit splitting for large systems**

In this section we demonstrate that the splittings corresponding to the quantum circuit ansatz are zero for sufficiently large systems; however, we note that the QC states are a good starting points for the DMRG-X calculation regardless.

The eigenstates that the quantum circuit approximation of tLIOMs predicts are given by

\[
|\psi_{\pm}\rangle = U_{SR}|t_{\pm}\rangle,
\]

where \( |t_{\pm}\rangle \) is an eigenstate of the topological stabilizers \( S_j^{\text{topo}} = i\gamma_{2j-1}\gamma_{2j+1} \) with eigenvalues \( t_j = \pm 1 \) and a \( \pm 1 \).
eigenstate of the bi-local stabilizer $S^{nl} = i\gamma_2 N \gamma_1$. It may be written in terms of stabilizers as
\[
|t_\pm\rangle\langle t_\pm| = \left(\frac{1}{2} \pm \frac{1}{2} S^{nl}\right) \prod_{j=1}^{N-1} \left(\frac{1}{2} + \frac{1}{2} t_j S^{topo}_j\right).
\]  
(C2)

Using this representation, the expectation value for the energy can be expressed as
\[
\langle \psi_\pm | H | \psi_\pm \rangle = \text{Tr} \left[ \left(\frac{1}{2} \pm \frac{1}{2} S^{nl}\right) \prod_{j=1}^{N-1} \left(\frac{1}{2} + \frac{1}{2} t_j S^{topo}_j\right) U_{SR}^\dagger H U_{SR} \right],
\]  
(C3)

and the corresponding splitting is given by
\[
\Delta E = \langle \psi_+ | H | \psi_+ \rangle - \langle \psi_- | H | \psi_- \rangle = \text{Tr} \left[ S^{nl} \prod_{j=1}^{N-1} \left(\frac{1}{2} + \frac{1}{2} t_j S^{topo}_j\right) U_{SR}^\dagger H U_{SR} \right].
\]  
(C4)

To further analyze the expression, we can expand the Hamiltonian in local terms
\[
H = \sum_{k=1}^{N} h_k;
\]  
(C5)

with that the splitting is given by
\[
\Delta E = \sum_{k=1}^{N} \text{Tr} \left[ S^{nl} \prod_{j=1}^{N-1} \left(\frac{1}{2} + \frac{1}{2} t_j S^{topo}_j\right) U_{SR}^\dagger h_k U_{SR} \right].
\]  
(C6)

Since the quantum circuit acts locally, the causal cones [31] of each term $h_k$ in the Hamiltonian will, for a sufficiently large system size $N$, not cover both of the Majorana modes making up $S^{nl}$. Hence, either $\gamma_1$ or $\gamma_N$ (or both) will factor from the trace, leading to $\Delta E = 0$.

The true splitting is of the order $\exp(-N/\xi_L)$, i.e., it is due to the exponentially decaying coupling between the end-modes in the bi-local integral of motion $\tau^{nl}$. It is thus not surprising that the QC approximation alone, which uses strictly short-range gates, cannot reproduce the splitting. We can capture $\Delta E$ better once the accuracy of the approximation has been improved, e.g., by feeding the QC approximate eigenstates into the DMRG-X algorithm. As pairs of QC initial states have the same configuration of approximate tLIOMs (other than for $\tau^{nl}$) and the DMRG-X algorithm tries to maintain a high overlap with the initial states, it will build up additional entanglement around these and thus improve their accuracy. This keeps the configuration of tLIOMs fixed for sufficiently large bond dimensions. Since the DMRG-X algorithm is based on matrix product states, it is able to build entanglement across the entire chain and will thus reproduce the energy splittings, including their dependence on system size, even for large systems.

Appendix D: Sensitivity to the thermal phase

To further illustrate the predictive power of the tLIOM approach, we have evaluated the (t)LIOMs along the $\sigma_{\nu}/\sigma_\mu = \tan(10^2)$ line, where we expect a smaller extent of the thermal phase (cf. Fig 1c). Along this line the figure of merit (Fig. 7a) has a smaller transition region in which both the LIOMs and tLIOMs poorly describe the system, consistent with a larger regime being a part of an MBL phase. Furthermore, focussing on the entanglement entropy fluctuations (Fig. 7b), we find that the maxima they assume are smaller. Since the thermal phase is responsible for generating those maxima, this is also consistent with a smaller extent.

[1] L. Fleishman and P. W. Anderson, Phys. Rev. B 21, 2366 (1980)
[2] I. V. Gornyi, A. D. Mirlin, and D. G. Polyakov, Phys. Rev. Lett. 95, 206603 (2005)
[3] D. M. Basko, I. L. Aleiner, and B. L. Altshuler, Ann. Phys. 321, 1126 (2006)
[4] M. Znidaric, T. Prosen, and P. Prelovšek, Phys. Rev. B 77, 064426 (2008)
