Classical Simulation of Short-Time Quantum Dynamics

Dominik S. Wild[1,*] and Álvaro M. Alhambra[1,2,†]

[1]*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, Garching D-85748, Germany*
[2]*Instituto de Física Teórica Universidad Autónoma de Madrid-Consejo Superior de Investigaciones Científicas, C/ Nicolás Cabrera 13-15, Cantoblanco, Madrid 28049, Spain*

Recent progress in the development of quantum technologies has enabled the direct investigation of the dynamics of increasingly complex quantum many-body systems. This motivates the study of the complexity of classical algorithms for this problem in order to benchmark quantum simulators and to delineate the regime of quantum advantage. Here, we present classical algorithms for approximating the dynamics of local observables and nonlocal quantities such as the Loschmidt echo, where the evolution is governed by a local Hamiltonian. For short times, their computational cost scales polynomially with the system size and the inverse of the approximation error. In the case of local observables, the proposed algorithm has a better dependence on the approximation error than algorithms based on the Lieb-Robinson bound. Our results use cluster-expansion techniques adapted to the dynamical setting, for which we give a novel proof of their convergence. This has important physical consequences besides our efficient algorithms. In particular, we establish a novel quantum speed limit, a bound on dynamical phase transitions, and a concentration bound for product states evolved for short times.

I. INTRODUCTION

The study of the dynamics of quantum many-body models is a highly active area of research in quantum information science, both from the perspective of physics and of computation. Probing dynamics provides access to a wealth of physical phenomena and can enable the solution of hard computational problems. Existing quantum simulators are already allowing us to explore quantum dynamics of high complexity. To assess the potential for quantum speed-up, it is important to understand the reach of classical methods for these tasks. Unless all quantum computations can be classically simulated, i.e., BPP = BQP, classical algorithms will be unable to approximate quantum dynamics in an arbitrary setting. Nevertheless, there exist restricted regimes in which efficient classical simulation is possible. An important special case is evolution for short times, during which the quantum information will not spread much. Tensor-network methods illustrate this point, as they provide provably efficient means of simulating unitary dynamics in one spatial dimension for short times [1,2].

In this work, we characterize the computational complexity of short-time dynamics under a local Hamiltonian more generally. Locality in this context means that the Hamiltonian can be written as a sum of operators supported on small subsystems. We do not require geometric locality and we do not restrict the Hamiltonian to a finite-dimensional lattice. Instead, we only impose that every term in the Hamiltonian overlaps with a constant number of other terms. These conditions are satisfied by a wide class of physically relevant Hamiltonians, including parent Hamiltonians of quantum low-density parity-check (LDPC) error-correcting codes [3]. Dynamics governed by Hamiltonians of this type are amenable to *cluster expansions*, which have long been used in both classical and quantum statistical mechanics of lattice models [4–8], leading to results such as the uniqueness of Gibbs states [9,10], efficient approximation schemes for partition functions [11,12], the decay of correlations [13–15], and concentration bounds [16,17]. Despite their long history, cluster expansions have typically been applied to equilibrium properties, while dynamics have been rarely considered.

The paper is structured as follows. In the remainder of this section, we summarize the main results and give an overview of their implications for the complexity of quantum dynamics. In Sec. II, we define the cluster notation

*dominik.wild@mpq.mpg.de
†alvaro.alhambra@csic.es

used throughout. Section III discusses the results and the algorithm for local observables. In Sec. IV, we describe corresponding results for the Loschmidt echo. Physical consequences of these results are discussed in Sec. V. We conclude in Sec. VI with further remarks and open questions. The main text provides an overview of all the proof techniques, while technical details that are less crucial to the understanding are placed in appendixes.

A. Summary of results

Our first main result concerns the dynamics of a few-body observable A under a local Hamiltonian H.

Result 1 (informal version of Theorem 6). *Given a local Hamiltonian H, a few-body operator A, and a product state ρ, there exists an algorithm that approximates* $\mathrm{tr}(e^{iHt}Ae^{-iHt}\rho)$ *up to additive error ε with a run time of at most*

$$\mathrm{poly}\left[\left(\frac{1}{\varepsilon}e^{\pi|t|/t^*}\right)^{\exp(\pi|t|/t^*)}\right], \qquad (1)$$

where t^* is a positive constant.

The scaling with the time t is rather unfavorable but the computational cost is independent of the system size. Moreover, for any constant value of t, the run time has a polynomial dependence on $1/\varepsilon$, which is an improvement over all previously known algorithms, such as those based on Lieb-Robinson bounds. Cluster expansion can thus be seen as an alternative approach to analyzing the effective locality and light-cone structure of many-body dynamics.

Our second result characterizes the complexity of computing the Loschmidt echo, $\mathrm{tr}(e^{-itH}\rho)$.

Result 2 (informal version of Theorem 12). *Given a local Hamiltonian H, a product state ρ, and $|t| < t_L^*$, there exists an algorithm that approximates $\log\mathrm{tr}(e^{-iHt}\rho)$ up to additive error ε with a run time of at most*

$$n \times \mathrm{poly}\left[\left(\frac{n}{1-|t|/t_L^*}\frac{1}{\varepsilon}\right)^{1/\log(t_L^*/|t|)}\right], \qquad (2)$$

where t_L^* is a positive constant.

A key insight of this work is that when ρ is a product state, the objects analyzed in Results 1 and 2 fit the framework of cluster expansions that are commonly applied to the partition function $\mathrm{tr}(e^{-\beta H})$ [11,17,22]. We give a novel proof of the convergence of these expansions based on the counting of trees.

In both algorithms, the cluster expansions enable an efficient grouping of the terms of a Taylor series by clusters of

subsystems. Crucially, we show that only connected clusters contribute. Because the number of connected clusters grows at most exponentially with the size of the cluster, we can establish the convergence of the cluster expansion at short times by bounding the magnitude of the individual terms. The computational cost of the approximation algorithms follows by estimating the cost of computing a truncated cluster expansion while controlling the truncation error. For local observables, we are able to extend the algorithm beyond the radius of convergence of the cluster expansion using analytic continuation. The doubly exponential dependence on the evolution time t in Result 1 is a direct consequence of the analytic continuation scheme.

The above results have several important physics implications. Result 2 implies that dynamical phase transitions [23] cannot occur at times $t \le t_L^*$ for local Hamiltonians and product initial states. In addition, it establishes a novel quantum speed limit [24] that is independent of the system size, in stark contrast to previous results for general initial states [25,26]. A generalization of Result 2 to the multi-Hamiltonian Loschmidt echo $\mathrm{tr}(\rho \prod_l e^{-itH^{(l)}})$, with $H^{(l)}$ local Hamiltonians, allows us to prove Gaussian concentration bounds of local observables on states evolved for a short time, a case not covered by previous results [27–29]. More precisely, we show that the probability of measuring $H^{(2)}$ in the evolved product state $\rho(t) = e^{-itH^{(1)}}\rho e^{itH^{(1)}}$ away from the mean $\mathrm{tr}(\rho(t)H^{(2)})$ by δ is suppressed by $e^{\mathcal{O}(-\delta^2/n)}$.

B. Complexity of dynamics

Our main results have nontrivial consequences for the computational complexity of short-time quantum dynamics, which are summarized in Table I. For observables, it is known that approximating $\langle A(t)\rangle$ with additive error $\varepsilon = 1/\mathrm{poly}(n)$ up to times $t = \mathrm{poly}(n)$ is BQP-complete. This follows from the fact that determining the state of a single qubit at the output of a circuit with $\mathrm{poly}(n)$ gates is BQP-complete, combined with the existence of local Hamiltonians that simulate arbitrary quantum computations [18]. At the same time, Theorem 6 shows that we can compute $\langle A(t)\rangle$ classically with a similar error with computational cost $\mathrm{poly}(n)$ as long as $t = \mathcal{O}(1)$. This indicates a transition in the complexity of simulating local observables as the system evolves. The exact nature of this transition and the computational complexity of the intermediate regime $t \sim \mathrm{polylog}(n)$, where the cluster expansion fails to be efficient, remains an open problem.

For the Loschmidt echo, there are two meaningful notions of approximation. The first one is with a small multiplicative error, which is equivalent to a small additive error in the logarithm. For this, Theorem 12 shows that there is an efficient approximation algorithm for times $|t| < t_L^*$ with polynomial cost in both n and $1/\varepsilon$. Unlike in

TABLE I. The computational complexity of computing the quantities in the leftmost column with additive error $\varepsilon = 1/\text{poly}(n)$, where n is the system size, for different times. All expectation values are with respect to product initial states. The constants t^* and t_L^* are independent of the system size but depend on the details of the problem. Entries highlighted in blue are results from this work. Question marks denote regimes where the computational complexity is unknown.

	$t < t^*, t_L^*$	$t = \mathcal{O}(1)$	$t = \mathcal{O}(\text{polylog}(n))$	$t = \mathcal{O}(\text{poly}(n))$
$\langle A(t) \rangle$	P	P	?	BQP-complete [18]
$\log \langle e^{-iHt} \rangle$	P	#P-hard [19,20]	#P-hard	#P-hard
$\langle e^{-iHt} \rangle$	P	?	?	BQP-complete [21]

Theorem 6, it is not possible to extend this result to arbitrary times. If we take $\rho \propto \mathbb{I}$, $L(t)$ becomes an imaginary-time partition function. Approximating this for $|t| = \mathcal{O}(1)$ even with an $\mathcal{O}(1)$ multiplicative error has been shown to be #P-hard for 2-local classical Ising models [20]. Hence, there is only a constant gap between the times accessible with our algorithm and the #P-hard regime, where an efficient algorithm is unlikely to exist. The complexity of the analogous problem for real partition functions and the thermodynamic free energy has been recently considered in Ref. [30].

We may also consider the weaker additive approximation to the Loschmidt echo. Since $|L(t)| \leq 1$, Theorem 12 also implies that we can efficiently approximate the Loschmidt echo to an ε-additive error for $|t| < t_L^*$. Calculating the Loschmidt echo for circuits of polynomial size with additive error $\varepsilon = 1/\text{poly}(n)$ is BQP-complete [21]. To the best of our knowledge, the intermediate regime has not been explored.

II. SETUP

A. Hamiltonian

We consider a set of n spins, V. Each spin $v \in V$ is associated with a local Hilbert space $\mathcal{H}^v$, with $\dim \mathcal{H}^v = d$. The total Hilbert space is formed by the tensor product space $\mathcal{H} = \bigotimes_{v \in V} \mathcal{H}^v$. We call a subset of spins $X \subseteq V$ a subsystem. For any linear operator A on $\mathcal{H}$, we denote its support by $X = \text{supp}(A)$, i.e., X is the smallest subsystem in V on which A acts nontrivially.

Next, we formally define the notion of local Hamiltonians. Given a set of subsystems S, we write a Hamiltonian H as

$$ H = \sum_{X \in S} \lambda_X h_X, \qquad (3) $$

where each λ_X is a real coefficient and h_X is a Hermitian operator acting on the subsystem X such that $\text{supp}(h_X) = X$. The coefficients satisfy $|\lambda_X| \leq 1$ and are chosen such that $\|h_X\| = 1$, where $\|\cdot\|$ is the operator norm. A Hamiltonian is called k-local if it is a sum of terms that act on at most k sites or, equivalently, $|X| \leq k$ for all $X \in S$.

To characterize the connectivity of the Hamiltonian, we define the associated interaction graph G [22]. Given a set

of subsystems S, the interaction graph G is a simple graph with vertex set S. There is an edge between two vertices X and Y if the respective subsystems overlap. We denote the maximum degree of the interaction graph by $\mathfrak{d}$. Throughout this work, we only consider k-local Hamiltonians for which, in addition, $\mathfrak{d}$ is independent of the system size n. Each local term in the Hamiltonian therefore only overlaps with a constant number of other terms, which includes many physically relevant cases such as Hamiltonians with finite-range interactions. We point out that the number of terms $|S|$ in these Hamiltonians increases at most linearly with the number of spins n.

B. Clusters

We define a cluster as a nonempty multiset of subsystems from S. Here, multiset refers to a set with possibly repeated elements but without ordering. We use bold-font letters $\mathbf{V}, \mathbf{W}, \ldots$ to denote clusters. We call the number of times a subsystem X appears in a cluster $\mathbf{W}$ the multiplicity $\mu_{\mathbf{W}}(X)$. If X is not contained in $\mathbf{W}$, then $\mu_{\mathbf{W}}(X) = 0$. The size $|\mathbf{W}| = \sum_{X \in S} \mu_{\mathbf{W}}(X)$ of a cluster is the number of subsystems that it contains, including their multiplicity. The set of all clusters of size m is denoted by $\mathcal{C}_m$ and the set of all clusters by $\mathcal{C} = \bigcup_{m \geq 1} \mathcal{C}_m$.

We associate with every cluster $\mathbf{W}$ a simple graph $G_{\mathbf{W}}$, the so-called cluster graph. The vertices of $G_{\mathbf{W}}$ correspond to the subsystems in $\mathbf{W}$, with repeated subsystems also appearing as repeated vertices. Two distinct vertices X and Y are connected by an edge if and only if the respective subsystems overlap, i.e., $X \cap Y \neq \emptyset$. We say that a cluster $\mathbf{W}$ is connected if and only if $G_{\mathbf{W}}$ is connected. We use the notation $\mathcal{G}_m$ for the set of connected clusters of size m and $\mathcal{G} = \bigcup_{m \geq 1} \mathcal{G}_m$ for the set of all connected clusters. The following statement concerning the number of connected clusters is an essential ingredient of our algorithms.

Lemma 1 (Proposition 3.6 of Haah et al. [22]). *Given a subsystem $X \in S$, the number of clusters in $\mathcal{G}_m$ that contain X is bounded from above by $(e\mathfrak{d})^m$. Moreover, there exists a deterministic classical algorithm with run time $\exp(\mathcal{O}(m \log \mathfrak{d}))$ that outputs a list of all such clusters.*

The classical algorithm is given as Algorithm 1 in Haah et al. [22, Sec. 3.4].
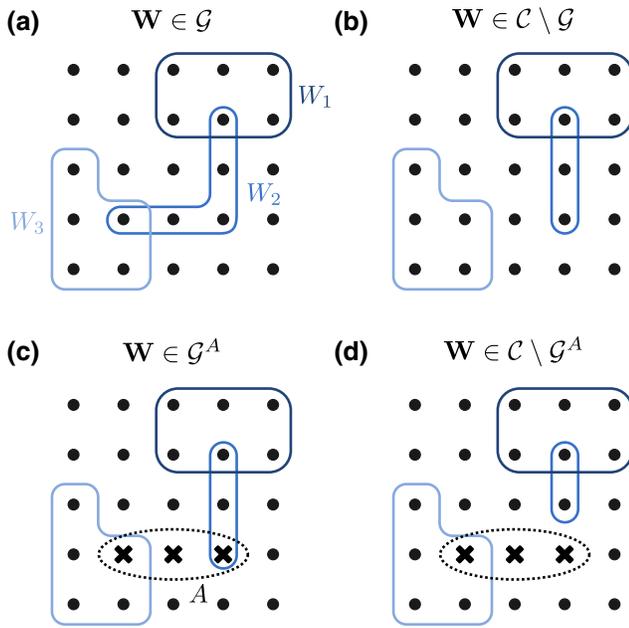
FIG. 1. A cluster of three subsystems $\mathbf{W} = \{W_1, W_2, W_3\}$ that is (a) connected, (b) disconnected, (c) completely connected to A, and (d) not completely connected to A. The black dots indicate individual spins and crosses highlight spins that form the support of A.

The union $\mathbf{W} = \mathbf{V}_1 \cup \mathbf{V}_2$ of two clusters $\mathbf{V}_1$ and $\mathbf{V}_2$ is defined as the union of the multisets, adding all multiplicities such that $\mu_{\mathbf{W}}(X) = \mu_{\mathbf{V}_1}(X) + \mu_{\mathbf{V}_2}(X)$ for all $X \in S$. Another set of clusters of special interest is formed by the clusters connected to the support of a few-body operator A. We say that a cluster $\mathbf{W}$ is completely connected to A if and only if the cluster graph $G_{\mathbf{W} \cup \{\mathrm{supp}(A)\}}$ is connected, where we assume for simplicity that A acts on a subsystem contained in the Hamiltonian such that $\mathrm{supp}(A) \in S$. We denote the set of such clusters of size m by $\mathcal{G}_m^A$ and $\mathcal{G}^A = \bigcup_{m \geq 1} \mathcal{G}_m^A$. The different sets of clusters are illustrated in Fig. 1.

Before proceeding, we introduce further notation related to clusters. It is sometimes convenient to assign a (nonunique) ordering to the subsystems in a cluster. For $\mathbf{W} \in \mathcal{C}_m$, we then write $\mathbf{W} = \{W_1, W_2, \ldots, W_m\}$. We make frequent use of the shorthands $\lambda^{\mathbf{W}} = \prod_{X \in S} \lambda_X^{\mu_{\mathbf{W}}(X)}$ and $\mathbf{W}! = \prod_{X \in S} \mu_{\mathbf{W}}(X)!$. We often take derivatives with respect to the parameters of the Hamiltonian for all subsystems contained in a cluster $\mathbf{W}$. To this end, we define the cluster derivative $\mathcal{D}_{\mathbf{W}}$, which acts on any function of the Hamiltonian parameters $\lambda = \{\lambda_X : X \in S\}$, as

$$\mathcal{D}_{\mathbf{W}} f(\lambda) = \left[\prod_{X \in S} \left(\frac{\partial}{\partial \lambda_X} \right)^{\mu_{\mathbf{W}}(X)} \right] f(\lambda) \Bigg|_{\lambda = 0}. \quad (4)$$

Here, the subscript $\lambda = 0$ means to set $\lambda_X = 0$ for all $X \in S$ after taking the derivatives. Hence, the cluster derivative isolates the contribution from the monomial $\lambda^{\mathbf{W}}$.

For any function $f(\lambda)$, we define its cluster expansion as the multivariate Taylor-series expansion in λ. With the above notation, the cluster expansion can be concisely written as

$$f(0) + \sum_{\mathbf{W} \in \mathcal{C}} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} f(\lambda). \quad (5)$$

Our goal is to establish conditions under which the cluster expansion converges to $f(\lambda)$ for different functions of interest.

We illustrate the above concepts by an example in Appendix B 1.

C. Cluster partitions

A partition P of a cluster $\mathbf{W}$ is a multiset of clusters $\{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{|P|}\}$ such that $\mathbf{W} = \mathbf{V}_1 \cup \mathbf{V}_2 \cup \cdots \cup \mathbf{V}_{|P|}$. We are particularly interested in partitions where every element is a connected cluster. We refer to these as partitions of $\mathbf{W}$ into connected subclusters. The set of all such partitions is denoted by $\mathcal{P}_c(\mathbf{W})$.

We introduce several quantities characterizing cluster partitions. We use tildes to distinguish these from similar quantities describing clusters. The multiplicity $\tilde{\mu}_P(\mathbf{V})$ is defined as the number of times the cluster $\mathbf{V}$ appears in the partition P. The size $|P| = \sum_{\mathbf{V} \in \mathcal{G}} \tilde{\mu}_P(\mathbf{V})$ is the number of clusters in P, including their multiplicities. We also use the shorthand

$$P! = \prod_{\mathbf{V} \in \mathcal{G}} \tilde{\mu}_P(\mathbf{V})! \quad (6)$$

For every partition $P \in \mathcal{P}_c(\mathbf{W})$, we define a simple graph $\tilde{G}_P$, called the partition graph of P. The vertices of $\tilde{G}_P$ are the clusters in P. Two clusters $\mathbf{V}, \mathbf{V}' \in P$ are connected by an edge if and only if they overlap, that is, there exist subsystems $X \in \mathbf{V}$ and $Y \in \mathbf{V}'$ such that $X \cap Y \neq \emptyset$. Alternatively, we may obtain $\tilde{G}_P$ from $G_{\mathbf{W}}$ as follows. For every $\mathbf{V} \in P$, we merge the corresponding vertices in $G_{\mathbf{W}}$ into a single vertex and remove all loops. If any of the remaining edges are repeated, they are reduced to a single edge. Figure 2 shows an example of a partition graph and illustrates its connection to the cluster graph.

III. LOCAL OBSERVABLES

A. Cluster expansion

We consider the expectation value of an observable A for an initial product state ρ evolving under a local

(a) $\mathbf{W} \in \mathcal{G}$ **(b)** W_1 $G_{\mathbf{W}}$

(c) $\tilde{G}_P$
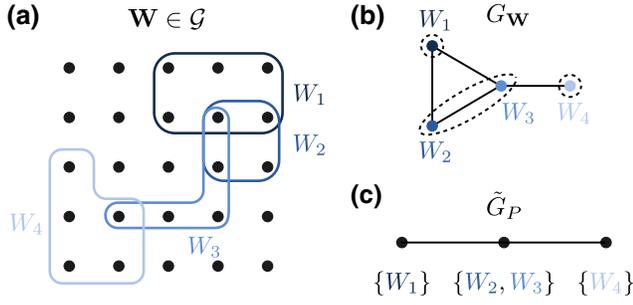
$\{W_1\}$ $\{W_2, W_3\}$ $\{W_4\}$

FIG. 2. An illustration of cluster and partition graphs. (a) A connected cluster $\mathbf{W} = \{W_1, W_2, W_3, W_4\}$ composed of four subsystems. The dots indicate individual spins. (b) The corresponding cluster graph $G_{\mathbf{W}}$. The dashed outlines show the partition $P = \{\{W_1\}, \{W_2, W_3\}, \{W_4\}\}$ of $\mathbf{W}$ into connected subclusters. (c) The partition graph $\tilde{G}_P$ associated with P.

Hamiltonian H. After time t, we have

$$\langle A(t) \rangle = \mathrm{tr}\left(e^{iHt} A e^{-iHt} \rho\right). \tag{7}$$

Due to the dependence of H on the parameter set λ, we may think of $\langle A(t) \rangle$ as a function of λ. The corresponding cluster expansion is given by

$$f_A(t) = \langle A(0) \rangle + \sum_{m=1}^{\infty} \sum_{\mathbf{W} \in \mathcal{C}_m} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \langle A(t) \rangle. \tag{8}$$

For $\mathbf{W} \in \mathcal{C}_m$, the cluster derivative can be explicitly evaluated as

$$\mathcal{D}_{\mathbf{W}} \langle A(t) \rangle = \frac{(it)^m}{m!}$$
$$\times \sum_{\sigma \in S_m} \mathrm{tr}\left([h_{W_{\sigma(1)}}, \cdots [h_{W_{\sigma(2)}}, \cdots [h_{W_{\sigma(m)}}, A]]]\rho\right), \tag{9}$$

where the sum runs over all permutations of the indices $\{1, 2, \ldots, m\}$.

For simplicity, we assume that the support of A is contained in the set of subsystems S of the local supports of the Hamiltonian, although this constraint can be readily relaxed. We establish the convergence of the cluster expansion for short times in two steps. First, we show that only clusters that are completely connected to A contribute to the sum.

Lemma 2. *For any cluster* $\mathbf{W} \notin \mathcal{G}^A$,

$$\mathcal{D}_{\mathbf{W}} \langle A(t) \rangle = 0. \tag{10}$$

Proof. Consider $\mathbf{W} \notin \mathcal{G}_m^A$ for any integer $m > 0$. For every $\sigma \in S_m$, there exists a positive integer $k \leq m$ such that the intersection of $W_{\sigma(k)}$ with $W_{\sigma(k+1)} \cup W_{\sigma(k+2)} \cup \cdots \cup$

$W_{\sigma(m)} \cup \mathrm{supp}(A)$ is empty. Then, $[h_{W_{\sigma(k)}}, \cdots [h_{W_{\sigma(k+1)}}, \cdots [h_{W_{\sigma(m)}}, A]]] = 0$ and the commutator in Eq. (9) vanishes. ∎

Second, we need to bound the magnitude of the cluster derivative when it does not vanish. Note that there are at most 2^m in the nested commutators of Eq. (9). Repeated application of the triangle inequality then yields $\left|\mathrm{tr}\left([h_{W_{\sigma(1)}}, \cdots [h_{W_{\sigma(2)}}, \cdots [h_{W_{\sigma(m)}}, A]]]\rho\right)\right| \leq 2^m \|A\|$. Hence, we find that for any $\mathbf{W} \in \mathcal{G}_m^A$,

$$|\mathcal{D}_{\mathbf{W}} \langle A(t) \rangle| \leq (2|t|)^m \|A\|. \tag{11}$$

By combining these observations, we are able to prove convergence of the cluster expansion for short times.

Proposition 3. *Consider an operator A for which* $\mathrm{supp}(A) \in S$ *and let* $|t| < t^* = 1/(2e\mathfrak{d})$. *Then,*

$$\left| \langle A(t) \rangle - \langle A(0) \rangle - \sum_{m=1}^{M} \sum_{\mathbf{W} \in \mathcal{G}_m^A} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \langle A(t) \rangle \right|$$
$$\leq e\mathfrak{d} \|A\| \frac{(|t|/t^*)^{M+1}}{1 - |t|/t^*}. \tag{12}$$

Proof. Consider the error of neglecting all clusters of size $m > M$ from the cluster expansion. Since $|\lambda_X| \leq 1$ for all $X \in S$, we can bound this error by

$$\left| \sum_{m=M+1}^{\infty} \sum_{\mathbf{W} \in \mathcal{G}_m^A} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \langle A(t) \rangle \right| \leq \sum_{m=M+1}^{\infty} (2|t|)^m |\mathcal{G}_m^A| \|A\|. \tag{13}$$

By assumption, $\mathrm{supp}(A) \in S$, which allows us to obtain every cluster in $\mathcal{G}_m^A$ by starting from a cluster $\mathbf{W} \in \mathcal{G}_{m+1}$ that contains $X = \mathrm{supp}(A)$ and reducing the multiplicity $\mu_{\mathbf{W}}(X)$ by 1. It follows from Lemma 1 that $|\mathcal{G}_m^A| \leq (e\mathfrak{d})^{m+1}$. Substituting this bound into Eq. (13) yields a geometric series, which converges when $|t| < t^*$. Convergence of this series implies the convergence of the cluster expansion such that $f_A(t) = \langle A(t) \rangle$ and leads to the error bound in the lemma. ∎

We highlight that Proposition 3 holds for any quantum state ρ. The restriction to product states only becomes relevant when bounding the computational cost. In addition, the proposition is valid for complex values of t.

B. Computation for short times

We now discuss the computational cost of estimating $\langle A(t) \rangle$ for a time $|t| < t^*$ up to additive error $\varepsilon \|A\|$. For simplicity, we only give the asymptotic scaling of the algorithm with ε and $|t|/t^*$ and suppress the dependence

on constant parameters such as the locality k or the connectivity $\mathfrak{d}$ of the Hamiltonian. Moreover, we exclude issues of finite numerical precision, which have been addressed rigorously by Haah *et al.* [22], from our considerations.

Our approximation algorithm computes the cluster expansion for all clusters up to size M, the value of which is determined by ε and $|t|/t^*$. For all $m \leq M$, we proceed in two steps:

(i) Enumerate all the connected clusters in $\mathcal{G}_m^A$
(ii) Compute and add the contributions of every cluster

Step (i) can be carried out by using the algorithm in Lemma 1 to first compute clusters of size $m + 1$ containing $X = \text{supp}(A)$ and by subsequently reducing the multiplicity of X by one. The run time is $\exp(\mathcal{O}(m))$. The computational effort of step (ii) is dominated by the evaluation of the sum over nested commutators in Eq. (9). We show in Appendix A that this can be done for a product state in time $\exp(\mathcal{O}(m))$ by suitably grouping the $m!$ terms of the sum.

Together, the two steps lead to the following run time of the approximation algorithm.

Proposition 4. *Let ρ be a product state and $|t| < t^* = 1/(2e\mathfrak{d})$. There exists an algorithm that outputs an estimate $\hat{f}_A(t)$ with run time*

$$\text{poly}\left[\left(\frac{1}{1 - |t|/t^*}\frac{1}{\varepsilon}\right)^{1/\log(t^*/|t|)}\right] \quad (14)$$

such that $\left|\langle A(t)\rangle - \hat{f}_A(t)\right| \leq \varepsilon \|A\|$.

Proof. According to Proposition 3, truncating the cluster expansion at order $M > \log e\mathfrak{d}/(1 - |t|/t^*)\varepsilon / \log t^*/|t|$ leads to an error that is bounded from above by $\varepsilon\|A\|$. Following the above discussion of enumerating the clusters and computing their contributions, we see that the cluster expansion truncated at order M can be evaluated in time $\exp(\mathcal{O}(M))$. Picking the smallest integer M that guarantees the desired error bound yields Eq. (14). ∎

C. Computation for arbitrary times

The convergence result of Proposition 3 is independent of the system size n. Hence, $\langle A(t)\rangle$ remains analytic in the thermodynamic limit for all complex values of t that satisfy $|t| < t^*$. Given any $t_0 \in \mathbb{R}$, we may write $\langle A(t)\rangle = \text{tr}\left(e^{iH(t-t_0)}Ae^{-iH(t-t_0)}\rho'\right)$, where $\rho' = e^{-iHt_0}\rho e^{iHt_0}$ is another quantum state. This shows that $\langle A(t)\rangle$ is analytic on a disk in the complex plane of radius t^* around any point on the real axis. Stated differently, $\langle A(t)\rangle$ is analytic for all complex values of t satisfying $|\text{Im}(t)| < t^*$. This analytic
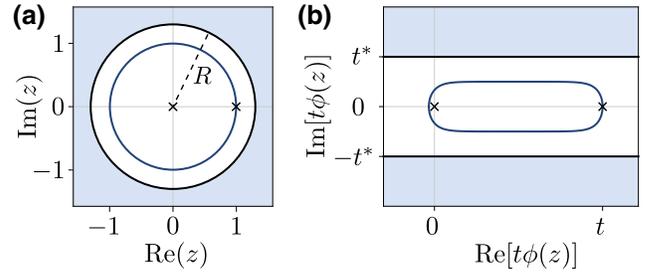
FIG. 3. An illustration of the analytic continuation scheme. (a) The function $\phi(z)$ is analytic in the domain $|z| \leq R$. The blue circle is the unit circle. (b) The function $\langle A(t\phi(z))\rangle$ is analytic for all z such that $|\text{Im}[t\phi(z)]| < t^*$. Given $t \in \mathbb{R}$, $t\phi(z)$ maps the unit circle to the elongated shape outlined by the blue curve.

structure provides a strategy to compute $\langle A(t)\rangle$ for a product state ρ for all $t \in \mathbb{R}$ and any system size n by means of analytic continuation.

While there are many approaches to analytic continuation, we pick here a concrete method that employs a function $\phi(z)$ that maps a disk onto an elongated region along the real axis. For some $R > 1$ and $w > 0$, we assume that $\phi(z)$ satisfies the following three properties:

(i) $\phi(z)$ is analytic on the closed disk $D_R = \{z \in \mathbb{C} : |z| \leq R\}$.
(ii) $\phi(0) = 0$ and $\phi(1) = 1$.
(iii) $|\text{Im}(\phi(z))| \leq w$ for all $z \in D_R$.

We show below, using an explicit example, that such a function exists.

Next, we define $f(z) = \langle A(t\phi(z))\rangle$, where $t \in \mathbb{R}$ is the time at which we want to evaluate $\langle A(t)\rangle$. It follows from property (ii) that $f(1) = \langle A(t)\rangle$. Because $\langle A(t)\rangle$ is analytic when $|\text{Im}(t)| < t^*$, properties (i) and (iii) together ensure that $f(z)$ is analytic on D_R, provided that $w|t| < t^*$. These relations are illustrated in Fig. 3.

Next, we compute the Taylor series of $f(z)$ at $z = 0$ up to order M. With $\langle A(t)\rangle = \sum_{l=0}^\infty A_l t^l$ and $\phi(z) = \sum_{l=0}^\infty \phi_l z^l$, we have

$$\frac{1}{k!}\frac{d^k f(z)}{dz^k}\bigg|_{z=0} = \sum_{l=1}^k A_l t^l \sum_{\substack{m_1,\ldots,m_l \geq 1 \\ m_1 + \cdots + m_l = k}} \phi_{m_1}\cdots\phi_{m_l}, \quad (15)$$

where we use the fact that $\phi_0 = 0$. We can obtain $A_l t^l$ from the cluster expansion as $A_l t^l = \sum_{\mathbf{W}\in\mathcal{G}_l}\lambda^{\mathbf{W}}\mathcal{D}_{\mathbf{W}}\langle A(t)\rangle/\mathbf{W}!$, which takes time $\exp(\mathcal{O}(l))$ to evaluate. The sum over the Taylor coefficients of $\phi(z)$ involves $\binom{l-1}{k-1} \leq 2^{l-1}$ terms. Assuming that the individual coefficients ϕ_l can be computed in time $\exp(\mathcal{O}(l))$, it thus takes time $\exp(\mathcal{O}(k))$ to compute the kth Taylor coefficient of $f(z)$, and time $\exp(\mathcal{O}(M))$ to compute the full Taylor series up to order M.

To bound the truncation error of the Taylor series, we make use of the following standard result in complex analysis (see, e.g., Ref. [12, Proposition 18]).

Lemma 5. *We denote by D_R the closed disk of radius R centered at $z = 0$ in the complex plane, i.e., $D_R = \{z \in \mathbb{C} | |z| \leq R\}$. Let $f(z)$ be a complex function that is bounded by $|f(z)| \leq F$ and analytic for all $z \in D_R$. Given $\alpha < 1$, for all $z \in D_{\alpha R}$, the error of approximating $f(z)$ by a truncated Taylor series of order M is bounded by*

$$\varepsilon_M(z) = \left| f(z) - \sum_{m=0}^{M} \frac{1}{m!} f^{(m)}(0) z^m \right| \leq \frac{\alpha^{M+1}}{1-\alpha} F. \quad (16)$$

Combining this lemma with the above considerations for computing the Taylor series yields an algorithm to estimate $\langle A(t) \rangle$ for all $t \in \mathbb{R}$.

Theorem 6. *Given $t > 0$, there is an algorithm that outputs the estimate $\hat{f}_A(t)$ with run time*

$$\text{poly} \left[\left(\frac{1}{\varepsilon} e^{\pi t / t^*} \right)^{\exp(\pi t / t^*)} \right] \quad (17)$$

such that for any product state ρ,

$$\left| \langle A(t) \rangle - \hat{f}_A(t) \right| \leq \varepsilon \|A\|. \quad (18)$$

Proof. We bound the truncation error of the Taylor series of $f(z)$ at $z = 1$ by letting $\alpha = 1/R$ in Lemma 5. This yields

$$\varepsilon_M(1) \leq \frac{1}{R^M (R-1)} \max_{z \in D_R} |f(z)|. \quad (19)$$

By the same argument that we use to show that $\langle A(z) \rangle$ is analytic for all $|\text{Im}(z)| < t^*$, Proposition 3 implies that

$$\langle A(z) \rangle \leq \frac{e \mathfrak{d}}{1 - |\text{Im}(z)|/t^*} \|A\| \quad (20)$$

for all z satisfying $|\text{Im}(z)| < t^*$. Since $|\text{Im}(\phi(z))| \leq w$, assuming $wt/t^* < 1$, we have

$$\max_{z \in D_R} |f(z)| \leq \frac{e \mathfrak{d}}{1 - wt/t^*} \|A\|. \quad (21)$$

Hence,

$$\varepsilon_M(1) \leq \frac{e \mathfrak{d}}{1 - wt/t^*} \frac{1}{R^M (R-1)} \|A\|. \quad (22)$$

The parameters R and w cannot be chosen entirely at will owing to the constraints on $\phi(z)$. We do not attempt to

address this issue in general but instead consider the concrete example $\phi(z) = \log(1 - z/R') / \log(1 - 1/R')$ with $R' > R > 1$, inspired by Ref. [31, Lemma 2.2.3]. This function clearly satisfies the requirements (i) and (ii). For requirement (iii), we assume that the branch of the logarithm is chosen such that $|\text{Im}(\phi(z))| \leq -\pi/[2 \log(1 - 1/R')]$. We therefore set $w = -\pi/[2 \log(1 - 1/R')]$. Recalling that $w < t^*/t$, we separately let $wt/t^* = \eta$ for some $\eta < 1$. We can always choose $R < R'$ such that $R'^m(R' - 1) = 2R^M(R - 1)$, allowing us to replace R by R' in Eq. (22) at the cost of a factor of 2. For the particular choice of $\phi(z)$, we thus obtain

$$\varepsilon_M(1) \leq \frac{2e \mathfrak{d}}{1 - \eta} \left(1 - e^{-\pi t / 2 \eta t^*} \right)^M \left(e^{\pi t / 2 \eta t^*} - 1 \right) \|A\|. \quad (23)$$

It follows from this expression that truncating at order $M > e^{\pi t / 2 \eta t^*} \log \left(2e \mathfrak{d}/1 - \eta 1/\varepsilon e^{\pi t / 2 \eta t^*} \right)$ guarantees an error $\varepsilon_M(1) \leq \varepsilon \|A\|$. Thus, our output $\hat{f}_A(t)$ is the Taylor series of $f(z)$ at $z = 1$ up to the smallest integer satisfying the lower bound on M. Since the computational cost is exponential in M, the theorem follows by setting $\eta = 1/2$. ∎

The above approach yields an algorithm to estimate $\langle A(t) \rangle$ with a computational cost that scales polynomially with $1/\varepsilon$ for any fixed real time t. The cost, however, has a doubly exponential dependence on t/t^*, which renders this approach unsuitable for practical computations at long times. The chain of disks, another common method of analytic continuation, also yields a doubly exponential dependence of the computational cost on the time t [32]. We conjecture that this scaling is an unavoidable consequence of analytic continuation. More efficient continuation algorithms may be available if the analytic domain of $\langle A(t) \rangle$ extends along the imaginary direction beyond a constant. This is, however, not possible in general because there exist local observables and Hamiltonians for which $\langle A(t) \rangle$ becomes nonanalytic in the thermodynamic limit at a constant imaginary time [33]. This also indicates that our convergence result is optimal up to an improvement of t^* by a constant factor. We remark that the above procedure can be adapted to not only yield expectation values with product initial states but also an M-local approximation of the operator $A(t)$.

D. Comparison with the Lieb-Robinson bound

The Lieb-Robinson bound [34] offers an alternative method for computing the time evolution of local operators. It implies that

$$\left| \langle A(t) \rangle - \langle e^{iH_l t} A e^{-iH_l t} \rangle \right| \leq C e^{(vt-l)/\xi} \|A\|, \quad (24)$$

where C and ξ are constants and v is the Lieb-Robinson velocity. The Hamiltonian $H_l = \sum_{X \in S: \, \text{dist}(X,A) \leq l} \lambda_X h_X$

contains all local terms within the graph distance $\text{dist}(X, A)$ between the operators X and A on the interaction graph G.

To compute $\langle A(t) \rangle$ to within additive error $\varepsilon \|A\|$, it suffices to compute $\langle e^{-iH_l t} A e^{iH_l t} \rangle$ on a region of radius $l = vt + \xi \log(C/\varepsilon)$. In a lattice in D dimensions, the computational cost of performing exact diagonalization on this region is exponential in $[vt + \xi \log(C/\varepsilon)]^D = \xi^D \log^D \left[e^{vt/\xi} (C/\varepsilon) \right]$. For $D > 1$, this yields an algorithm the cost of which is superpolynomial in both $e^{vt/\xi}$ and $1/\varepsilon$, as opposed to the polynomial dependence on $1/\varepsilon$ in Theorem 6. The difference is even more marked in expander graphs, for which the number of sites at a distance l grows as $\exp(\mathcal{O}(l))$. The Lieb-Robinson method has a run time that is exponential in $\text{poly}(1/\varepsilon)$ and doubly exponential in t for such graphs.

IV. LOSCHMIDT ECHO

A. Cluster expansion

We now focus on the Loschmidt echo

$$L(t) \equiv \text{tr}\left(e^{-iHt} \rho \right), \quad (25)$$

where ρ is a product state on the n qubits. This is an important quantity in the study of dynamics of quantum systems. It appears in diverse contexts such as quantum chaos [35], as the characteristic function of the local density of states [36], and in algorithms for quantum simulation [37]. As we discuss below, it is also a key quantity in the description of dynamical phase transitions and other relevant phenomena.

We consider the logarithm $\log L(t)$, the multivariate Taylor expansion of which can be written in terms of cluster derivatives as

$$f_L(t) = \sum_{m=1}^{\infty} \sum_{\mathbf{W} \in \mathcal{C}_m} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(t). \quad (26)$$

Our goal is to establish sufficient criteria for the convergence of this expansion. Working with the logarithm of $L(t)$ greatly reduces the number of clusters involved, since only connected ones contribute.

Lemma 7. *If ρ is a product state, then for any disconnected cluster* $\mathbf{W}$,

$$\mathcal{D}_{\mathbf{W}} \log L(t) = 0. \quad (27)$$

Proof. Since $\mathbf{W} \notin \mathcal{G}$, there exists a decomposition $\mathbf{W} = \mathbf{W}_A \cup \mathbf{W}_B$ such that the cluster graphs $G_{\mathbf{W}_A}$ and $G_{\mathbf{W}_B}$ are disconnected components of $G_{\mathbf{W}}$. We define $H_A = \sum_{X \in \mathbf{W}_A} \lambda_X h_X$ and $H_B = \sum_{X \in \mathbf{W}_B} \lambda_X h_X$, where each subsystem is included at most once in the sum, even if it appears with higher multiplicity in the cluster. Clearly,

$\text{supp}(H_A) \cap \text{supp}(H_B) = \emptyset$, which implies $[H_A, H_B] = 0$ and

$$\mathcal{D}_{\mathbf{W}} \log L(t) = \mathcal{D}_{\mathbf{W}} \log \text{tr}\left(e^{-iH_A t} e^{-iH_B t} \rho \right). \quad (28)$$

For ρ a product state, we further have

$$\text{tr}\left(e^{-iH_A t} e^{-iH_B t} \rho \right) = \text{tr}\left(e^{-iH_A t} \rho \right) \text{tr}\left(e^{-iH_B t} \rho \right) \quad (29)$$

and thus

$$\mathcal{D}_{\mathbf{W}} \log L(t)$$
$$= \mathcal{D}_{\mathbf{W}} \log \text{tr}\left(e^{-iH_A t} \rho \right) + \mathcal{D}_{\mathbf{W}} \log \text{tr}\left(e^{-iH_B t} \rho \right). \quad (30)$$

The first term vanishes because H_A is independent of λ_X for any $X \in \mathbf{W}_B$. Similarly, the second term is independent of λ_X for $X \in \mathbf{W}_A$. ∎

Hence, we can restrict the sum in Eq. (26) to connected clusters.

We next bound the cluster derivative using the bounded connectivity of the cluster and partition graphs. Many properties of a graph G are captured by the Tutte polynomial $T_G(x, y)$ (see, e.g., Ref. [38]). For instance, $T_G(1, 1)$ is the number of spanning trees (or spanning forests if the graph is not connected), $T_G(2, 1)$ is the number of forests, and $T_G(2, 2)$ is $2^{|E|}$, where $|E|$ is the number of edges.

Using the Tutte polynomial, the cluster derivative of $\log L(t)$ can be concisely expressed as in the following lemma.

Lemma 8. *For any $\mathbf{W} \in \mathcal{G}_m$, the cluster derivative of $\log L(t)$ can be written as*

$$\mathcal{D}_{\mathbf{W}} \log L(t) = (-it)^m$$
$$\times \sum_{P \in \mathcal{P}_c(\mathbf{W})} (-1)^{|P|-1} N_P(\mathbf{W}) T_{\tilde{G}_P}(1, 0) \prod_{\mathbf{V} \in P} \langle h^{\mathbf{V}} \rangle_s, \quad (31)$$

where we introduce the symmetrized expectation value

$$\langle h^{\mathbf{V}} \rangle_s = \frac{1}{|\mathbf{V}|!} \sum_{\sigma \in S_{|\mathbf{V}|}} \text{tr}\left(h_{V_{\sigma(1)}} h_{V_{\sigma(2)}} \cdots h_{V_{\sigma(|\mathbf{V}|)}} \rho \right) \quad (32)$$

and the combinatorial factor

$$N_P(\mathbf{W}) = \frac{\mathbf{W}!}{P! \prod_{\mathbf{V} \in P} \mathbf{V}!}. \quad (33)$$

We prove this lemma in Appendix B 2. A similar statement has been reported by Mann and Helmuth [11], with Helmuth, Perkins, and Regts [39] pointing out the relation to the Tutte polynomial.

We make use of Lemma 8 to derive the following upper bound on the cluster derivative.

Proposition 9. *Let* $\mathbf{W} \in \mathcal{G}_m$. *Then,*

$$\left| \frac{1}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(t) \right| \leq [2e(\mathfrak{d}+1)|t|]^m. \qquad (34)$$

We defer the proof to Appendix B 4. In rough terms, it proceeds by observing that $|\langle h^{\mathbf{V}} \rangle_s| \leq 1$ and $0 \leq T_{\tilde{G}_P}(1,0) \leq T_{\tilde{G}_P}(1,1)$, where $T_{\tilde{G}_P}(1,1)$ is equal to the number of spanning trees of $\tilde{G}_P$. The sum of these trees over all partitions can then be bounded by the number of spanning trees of the original cluster graph $G_{\mathbf{W}}$. This number is smaller than an exponential in m times $\mathbf{W}!$, yielding Eq. (34).

Having bounded each term in the cluster expansion, it remains to bound the number of terms, i.e., the number of connected clusters of size m. This is done with Lemma 1 and the fact that there are $|S|$ subsystems on the lattice such that the total number is bounded by $|S|(e\mathfrak{d})^m$.

The main result of this section is the following theorem.

Theorem 10. *The logarithm of the Loschmidt echo,* $\log L(t)$, *is analytic for* $|t| < t_L^* = 1/[2e^2\mathfrak{d}(\mathfrak{d}+1)]$ *and the truncation error of the cluster expansion can be bounded by*

$$\left| \log L(t) - \sum_{m=1}^{M} \sum_{\mathbf{W} \in \mathcal{G}_m} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(t) \right|$$

$$\leq |S| \frac{(|t|/t_L^*)^{M+1}}{1 - |t|/t_L^*}. \qquad (35)$$

Proof. Lemma 7 and Proposition 9, together with the bound on the number of clusters, imply that

$$\left| \sum_{\mathbf{W} \in \mathcal{G}_m} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(t) \right| \leq |S| [2e^2\mathfrak{d}(\mathfrak{d}+1)|t|]^m. \qquad (36)$$

The result then follows by considering the weight of the terms of the Taylor expansion with $m > M$. ∎

The overall argument of this section mirrors the steps of previous results on Gibbs states [17,22]. Our new technical contributions are the expression for the cluster derivative in Lemma 8 and the bound in Proposition 9, for which we use a novel proof strategy based on counting trees. We note that convergence results similar to Theorem 10 can also be proved using the general framework of abstract polymer models [8,10,11].

B. Computation of the Loschmidt echo

The above Taylor approximation allows for an efficient classical estimation of the Loschmidt echo for short times.

Theorem 10 guarantees that to approximate $\log L(t)$, it suffices to calculate the terms in the series up to some order M. Recall that the mth-order term is given by

$$\sum_{\mathbf{W} \in \mathcal{G}_m} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(t). \qquad (37)$$

As in Sec. III B, we need two steps to calculate these: (i) enumerate all the connected clusters in $\mathcal{G}_m$ and (ii) compute and sum the cluster derivative for every connected cluster according to Eq. (37). Lemma 1 addresses the first step. For the second one, we need to bound the cost of computing cluster derivatives. Related bounds have previously been stated in several works [11,22,40], where the computation has proceeded either by directly differentiating $\log L(t)$ or by summing the terms of an expansion similar to that in Lemma 8. Here, we pursue the latter approach, stating the computational cost in the following proposition. As in Sec. III, we ignore complications arising due to finite numerical precision.

Proposition 11. *There exists a deterministic algorithm that outputs* $\mathcal{D}_{\mathbf{W}} \log L(t)$ *for* $\mathbf{W} \in \mathcal{G}_m$ *with run time* $\exp(\mathcal{O}(m))$.

Proof. The algorithm evaluates the expression in Lemma 8. There are three nontrivial contributions to the run time:

(i) Enumerating the partitions of $\mathbf{W}$ into connected subclusters. This takes time $\exp(\mathcal{O}(m))$ by the following algorithm. Assign to each subsystem in $\mathbf{W}$ a unique label from the set $\{1, 2, \dots, m\}$. List all compositions of m, i.e., ordered tuples of positive integers $(n_1, n_2, \dots, n_l)$ such that their sum equals m. There are 2^{m-1} distinct compositions, which can be enumerated in time $\exp(\mathcal{O}(m))$. For each composition, find all connected clusters $\mathbf{V}_1$ of size n_1 that are contained in $\mathbf{W}$ and include the subsystem labeled 1. By Lemma 1, this step can be carried out in time $\exp(\mathcal{O}(n_1))$ by enumerating all clusters connected to subsystem 1 and removing the ones that are not contained in $\mathbf{W}$. Next, for each $\mathbf{V}_1$, find all connected clusters $\mathbf{V}_2$ of size n_2 that are contained in $\mathbf{W} \setminus \mathbf{V}_1$ and include the subsystem with the smallest label remaining in $\mathbf{W} \setminus \mathbf{V}_1$. This step takes a computational time $\exp(\mathcal{O}(n_2))$. We iterate this procedure, removing the new cluster $\mathbf{V}_i$ from the original cluster in every step until $\mathbf{W} = \bigcup_{i=1}^{l} \mathbf{V}_i$. The procedure takes time $\exp(\mathcal{O}(n_1)) \exp(\mathcal{O}(n_2)) \cdots \exp(\mathcal{O}(n_l)) = \exp(\mathcal{O}(m))$. The above steps produce a list of length $\exp(\mathcal{O}(m))$, which includes all desired partitions. Duplicates may appear, although these can be removed in time $\exp(\mathcal{O}(m))$.

(ii) Computing the Tutte polynomial $T_{\tilde{G}_P}(1,0)$, where $\tilde{G}_P$ has $|P| \leq m$ vertices. This can be done in time $\exp(\mathcal{O}(|P|))$ using the algorithm by Björklund *et al.* [41].

(iii) Computing the symmetrized expectation value $\langle h^{\mathbf{V}} \rangle_s$ for $\mathbf{V} \in \mathcal{G}_l$ with $l \leq m$. The approach in Appendix A—Eq. (A2) in particular—can be adapted to carry out this computation in time $\exp(\mathcal{O}(l))$.

■

With these ingredients, the main result is as follows.

Theorem 12. *For times* $|t| < t_L^* = 1/[2e^2\mathfrak{d}(\mathfrak{d}+1)]$*, there exists a classical algorithm with run time*

$$|S| \times \text{poly}\left[\left(\frac{|S|}{1-|t|/t_L^*}\frac{1}{\varepsilon}\right)^{1/\log(t_L^*/|t|)}\right] \quad (38)$$

that outputs $\hat{f}_L(t)$ *such that* $|\log L(t) - \hat{f}_L(t)| \leq \varepsilon$.

Proof. Theorem 10 implies that the truncation error of the cluster expansion is smaller than ε when keeping terms up to $M > \log|S|/(1-|t|/t_L^*)\varepsilon/\log t_L^*/t$. With Lemma 1 and Proposition 11, this determines the run time in Eq. (38).

■

For a fixed $t < t_L^*$, the run time is polynomial in $1/\varepsilon$ as well as in the number of terms $|S|$, which is proportional to the system size n. The output approximates the Loschmidt echo by $\exp[\hat{f}_L(t)]$ with a multiplicative error

$$e^{-\varepsilon}\left|e^{\hat{f}_L(t)}\right| \leq |L(t)| \leq e^{\varepsilon}\left|e^{\hat{f}_L(t)}\right|. \quad (39)$$

Unlike in the case of local observables, it is in general not possible to analytically continue the Loschmidt echo beyond t_L^* because the zeros of $L(t)$, and thus any non-analyticities of $\log L(t)$, may be located anywhere in the complex plane, including the real axis.

C. Generalized Loschmidt echo

We now show that similar results hold for a Loschmidt echo with multiple Hamiltonians, defined as

$$L(t_1, t_2, \ldots, t_K) = L(\{t_l\}) = \text{tr}\left(\prod_{l=1}^{K} e^{-iH^{(l)}t_l}\rho\right). \quad (40)$$

We assume that each of the K Hamiltonians $\{H^{(l)}\}$ satisfies the same conditions as in Sec. II, so that they can all be written as

$$H^{(l)} = \sum_{X \in S} \lambda_X^{(l)} h_X^{(l)}. \quad (41)$$

We define the set of labeled subsystems $S^K = \{(X, l) : X \in S, l \in \{1, 2, \ldots K\}\}$, where the additional index keeps

track of the Hamiltonian. The corresponding interaction graph G^K may be viewed as K copies of the original interaction graph G, where vertices are connected if the subsystems overlap, independent of which copy they belong to. Note that if the maximum degree of the original interaction graph G was $\mathfrak{d}$, then the maximum degree of G^K is $K(\mathfrak{d}+1) - 1$.

A cluster $\mathbf{W}$ is now defined as a multiset of elements from S^K, with the set of clusters of size m denoted by $\mathcal{C}_m^K$ and the set of connected clusters by $\mathcal{G}_m^K$. The multiplicities of a subsystem and Hamiltonian label in a cluster are denoted by $\mu_{\mathbf{W}}((X, l))$. The cluster graph $G_{\mathbf{W}}^K$ is again constructed by connecting subsystems that overlap, independent of the Hamiltonian with which they are associated.

With this notation, the logarithm of the Loschmidt echo permits the cluster expansion

$$f_L(\{t_l\}) = \sum_{m=1}^{\infty} \sum_{\mathbf{W} \in \mathcal{G}_m^K} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(\{t_l\}). \quad (42)$$

Here, we introduce natural generalizations of our shorthands, $\lambda^{\mathbf{W}} = \prod_{(X,l) \in S^K} \left(\lambda_X^{(l)}\right)^{\mu_{\mathbf{W}}((X,l))}$, $\mathbf{W}! = \prod_{(X,l) \in S^K} \mu_{\mathbf{W}}((X,l))!$, and the cluster derivative

$$\mathcal{D}_{\mathbf{W}} f(\{t_l\}) = \left[\prod_{(X,l) \in S^K} \left(\frac{\partial}{\partial \lambda_X^{(l)}}\right)^{\mu_{\mathbf{W}}((X,l))}\right] f(\{t_l\})\bigg|_{\lambda=0}. \quad (43)$$

As we show in Appendix C, the results of Sec. IV A carry over directly, as long as we take into account the increased maximum degree of the interaction graph G^K. In particular, this means that for $\mathbf{W} \in \mathcal{G}_m^K$,

$$\left|\frac{1}{\mathbf{W}!}\mathcal{D}_{\mathbf{W}}\log L(\{t_l\})\right| \leq [2eK(\mathfrak{d}+1)]^m \prod_{l=1}^{K} |t_l|^{m_l}, \quad (44)$$

where m_l is the number of terms in $\mathbf{W}$ associated with $H^{(l)}$. Note the additional factor of K, which effectively reduces the threshold time by $1/K$.

The analysis of the computational cost is similar to that in Sec. IV B, as detailed in Appendix C. Hence, we obtain an analog of Theorems 10 and 12.

Theorem 13. *Let* $\tau = K\sum_{l=1}^{K} |t_l|/t_L^*$, *where* $t_L^* = 1/[2e^2\mathfrak{d}(\mathfrak{d}+1)]$. *The logarithm of the generalized Loschmidt echo,* $\log L(\{t_l\})$, *is analytic for* $\tau < 1$ *and its Taylor series*

converges as

$$\left| \log L(\{t_l\}) - \sum_{m=1}^{M} \sum_{\mathbf{W} \in \mathcal{G}_m^K} \prod_{l=1}^{K} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(\{t_l\}) \right|$$
$$\leq |S| \frac{\tau^{M+1}}{1-\tau}. \tag{45}$$

Moreover, there exists a classical algorithm with run time

$$|S| \times \text{poly} \left[\left(\frac{|S|}{1-\tau} \frac{1}{\varepsilon} \right)^{K/\log(1/\tau)} \right] \tag{46}$$

that outputs $\hat{f}_L(\{t_l\})$ such that $|\log L(\{t_l\}) - \hat{f}_L(\{t_l\})| \leq \varepsilon$.

We highlight again that the computational cost scales polynomially with the number of Hamiltonian terms $|S|$ and the inverse error $1/\varepsilon$. The number of Hamiltonians K enters in the shortened threshold time t_L^*/K and in the exponent in Eq. (46) as a computational overhead.

V. FURTHER IMPLICATIONS

Besides leading to efficient classical algorithms, the convergence of the cluster expansion has important physical consequences, which we discuss below.

A. Concentration bounds

Given a quantum state ρ, the outcome of a measurement of an observable A is a random variable. Assuming a projective measurement onto the eigenspaces of A, the probability of outcome x is given by $\Pr(x) = \text{tr}[\Pi(x)\rho]$, where $\Pi(x)$ is the projector onto the eigenspace of A with eigenvalue x. A concentration bound is an upper bound on the probability $\Pr[|x - \text{tr}(\rho A)| \geq \delta]$, i.e., the probability that x deviates from its mean $\text{tr}(\rho A)$ by more than δ. We focus on the Hamiltonian $H = \sum_X \lambda_X h_X$ as the observable. The energy distribution of a state plays an important role in thermalization and equilibration [42,43] and its concentration properties place limitations on the performance of variational quantum algorithms [29,44].

As a warm-up example, suppose that the Hamiltonian terms h_X act on distinct single sites and that ρ is a product state. Then, the measurement outcome of H is equal to the sum of the measurement outcomes of h_X, which are independent random variables. The Chernoff-Hoeffding bound for independent random variables implies that

$$\Pr[|x - \text{tr}(\rho H)| \geq \delta] \leq 2 \exp \left(-\frac{\delta^2}{2 \sum_X \|h_X\|^2} \right). \tag{47}$$

We note that the denominator in the exponent is proportional to the system size n such that deviations from the mean energy of order $\sqrt{n}$ are strongly suppressed.

This simple argument fails when the terms h_X in the Hamiltonian overlap or when ρ is not a product state because the outcomes of the measurements h_X are no longer independent. Nevertheless, similar bounds hold for local Hamiltonians and sufficiently weakly correlated states. A number of proof techniques have been employed to establish such results [27,28,45], including cluster expansion [40] (for results on large deviations, see also, e.g., Ref. [16]).

To illustrate the method, we use cluster expansion to give a concise proof of a concentration bound for the energy of product states, reproducing the main result of Ref. [27]. It follows from a standard argument (see, e.g., Ref. [40, Corollary 1]) that the probability of obtaining a measurement outcome x that is greater than the average energy by at least δ is bounded according to

$$\Pr[x - \text{tr}(\rho H) \geq \delta] \leq e^{-\delta \nu} \text{tr} \left[\rho \, e^{\nu(H - \text{tr}(\rho H))} \right]. \tag{48}$$

Next, we apply Theorem 10 with $t = i\nu$, where $0 \leq \nu < t_L^*$. Taking $M = 1$, the theorem implies that

$$\left| \log \text{tr}(\rho e^{\nu H}) - \nu \, \text{tr}(\rho H) \right| \leq |S| \frac{(\nu/t_L^*)^2}{1 - \nu/t_L^*}. \tag{49}$$

Substituting into Eq. (48) yields

$$\Pr[x - \text{tr}(\rho H) \geq \delta] \leq \exp \left[-\delta \nu + |S| \frac{(\nu/t_L^*)^2}{1 - \nu/t_L^*} \right]. \tag{50}$$

Choosing $\nu = \delta(t_L^*)^2/4|S|$ (which is always possible since $\delta/|S| \leq 1$ and $t_L^* < 1$) and applying the same bound for $\Pr(x - \text{tr}[\rho H] \leq \delta)$, we obtain

$$\Pr[|x - \text{tr}(\rho H)| \geq \delta] \leq 2e^{-(\delta t_L^*)^2/8|S|}. \tag{51}$$

We observe that this bound has the same dependence on δ and the system size as the bound for the special case in Eq. (47).

Theorem 13 enables us to extend the bound beyond product states. For short times, we can consider

$$\left| \log \text{tr} \left(e^{-itH^{(1)}} \rho e^{itH^{(1)}} e^{\nu H^{(2)}} \right) \right.$$
$$\left. - \nu \, \text{tr} \left(e^{-itH^{(1)}} \rho e^{itH^{(1)}} H^{(2)} \right) \right|, \tag{52}$$

from which we obtain the following concentration bound.

Corollary 14. *Let $H^{(1)}$ and $H^{(2)}$ be local Hamiltonians, let ρ be a product state, and let $|t| \leq t_L^*/7$. The probability that a projective measurement of the state $e^{-itH^{(1)}} \rho e^{itH^{(1)}}$ onto the eigenbasis of $H^{(2)}$ yields a value x that deviates from*

the expectation value $\mathrm{tr}[e^{-itH^{(1)}}\rho e^{itH^{(1)}}H^{(2)}]$ *by at least δ is bounded from above by*

$$\mathrm{Pr}\left[\left|x - \mathrm{tr}\left(e^{-itH^{(1)}}\rho e^{itH^{(1)}}H^{(2)}\right)\right| \geq \delta\right] \leq 2e^{\frac{-(\delta t_L^*)^2}{(250)^2|S|}}. \quad (53)$$

The proof is shown in Appendix D. The corollary shows that after evolving a product state for a short time under a local Hamiltonian, the energy distribution with respect to a (possibly different) local Hamiltonian is concentrated around the mean. Previous results along these lines cover many relevant cases but not time-evolved product states [17,27–29,45]. Since Theorem 13 works for any number of Hamiltonians, this corollary can also be extended to states of the form $e^{-itH^{(1)}}\cdots e^{-itH^{(K)}}\rho e^{itH^{(K)}}\cdots e^{itH^{(1)}}$, which often feature in variational quantum algorithms, for a correspondingly shorter threshold time.

B. Dynamical phase transitions

Dynamical phase transitions (DPTs) may be viewed as a real-time analog of thermal phase transitions [46]. The cluster expansion naturally constrains the time at which they can appear. Let us consider an infinite sequence of local Hamiltonians H_n on n particles under the assumptions of Sec. II and product states $|\Phi\rangle = \bigotimes_{i=1}^n |\phi\rangle_i$, such that $g_n(t) \equiv \log\langle\Phi|e^{-itH_n}|\Phi\rangle$. A DPT occurs when the following function is nonanalytic [23]:

$$G(t) = \lim_{n\to\infty} \frac{g_n(t)}{n}. \quad (54)$$

The following result is a consequence of Theorem 10.

Corollary 15. *$G(t)$ is analytic for $t < t_L^*$, and thus DPTs can occur only at later times.*

The proof is shown in Appendix E. The time scale t_L^* is hence a universal lower bound on the time at which dynamical phase transitions occur. Nonanalyticities can appear in the logarithm of the Loschmidt echo at times $t \sim \mathcal{O}(1)$, as can be seen from the simple case of noninteracting spins. This has also been demonstrated analytically for particular interacting models in one dimension [47]. This shows that t_L^* in Theorem 10 can be increased at most by a constant factor. Note also that Theorem 13 allows us to extend this corollary to some Floquet systems. The absence of dynamical phase transitions at short times is analogous to the fact that thermal phase transitions can only occur above some threshold inverse temperature β^* that depends on the details of the system.

C. Quantum speed limits

A quantum speed limit (QSL) is a bound on the time t_{QSL} that it takes for a state Φ evolving under a Hamiltonian H

to become orthogonal to itself. Formally,

$$t_{\mathrm{QSL}} = \min\{t : \langle\Phi|e^{-itH}|\Phi\rangle = 0\}. \quad (55)$$

The best-known general limits are the Mandelstam-Tamm and the Margolus-Levitin bounds [25,26]. When combined, they read

$$t_{\mathrm{QSL}} \geq \frac{\pi}{2}\max\left\{\frac{1}{\Delta H}, \frac{1}{\langle H\rangle}\right\}, \quad (56)$$

where $\langle H\rangle = \langle\Phi|H|\Phi\rangle$, $(\Delta H)^2 = \langle\Phi|(H - \langle H\rangle)^2|\Phi\rangle$ and we assume that all eigenvalues of H are positive. In many-body systems, however, we typically have that $\Delta H \sim n^{1/2}$ and $\langle H\rangle \sim n$, so that the bound vanishes with the system size. A simple consequence of Theorem 10 gives a significant improvement on the bound:

Corollary 16. *Let H be local and let $|\Phi\rangle$ be a product state. Then, $t_{\mathrm{QSL}} \geq t_L^*$.*

Proof. Theorem 10 shows that the logarithm of the fidelity is analytic for $t < t_L^*$. Since $\log(x)$ is nonanalytic at $x = 0$, this means that $|\langle\Phi|e^{-itH}|\Phi\rangle| > 0$ for $t < t_L^*$. ∎

Alternatively, the QSL also follows from the explicit lower bound in Eq. (39). By truncating the cluster expansion at order $M = 2$, we obtain the lower bound

$$|\langle\Phi|e^{-itH}|\Phi\rangle| \geq \exp\left[-|S|\frac{(|t|/t_L^*)^4}{1 - (|t|/t_L^*)^2}\right]e^{-\Delta H^2 t^2/2} \quad (57)$$

for all $|t| < t_L^*$. The dependence on $|t|/t_L^*$ in the first exponential is better than in Theorem 10 because all odd orders in the cluster expansion are purely imaginary and therefore do not contribute to the absolute value $|\exp[\hat{f}_L(t)]|$.

This result shows that the well-known QSLs of Eq. (56) do not give very tight bounds for product states evolving under local Hamiltonians. Let us remark that even if the fidelity does not become zero at early times, it does generically quickly become exponentially small in system size [48,49], which can also be seen from the upper bound in Eq. (39).

VI. SUMMARY AND OUTLOOK

We show that the cluster expansion of many dynamical quantities converges at short times, yielding efficient classical approximation algorithms as a by-product. We describe the implications for the complexity of quantum dynamics and discuss consequences for concentration bounds, which are linked to the performance of variational quantum algorithms [29,50], dynamical phase transitions, and quantum speed limits.

The proof strategy of our main results is based on counting the number of clusters that participate and bounding

their individual contributions to the sum. This last step diverges from established convergence proofs of cluster expansions in the literature on abstract polymer models [5, 8,10,51], which are based on iterative arguments. We follow more closely recent papers with results on Gibbs states [17,22,40], although we use an alternative expression for the cluster derivative involving the Tutte polynomial of the partition graph, which may be of independent interest.

Our work opens the door to many future research directions. It will be interesting to explore the optimality of our algorithms. For example, is it possible to improve the time dependence of Theorem 6 to the one given by the Lieb-Robinson bound (i.e., $e^{\mathcal{O}(t^D)}$ in D dimensions) while retaining the polynomial dependence of the inverse approximation error? Similarly, we may ask whether it is possible to extend the concentration bound, Corollary 14, to longer times. Related bounds on moments of the distribution have already been shown to hold for times up to $\mathcal{O}(\log n)$ [52]. One could also address in this context whether a sharp breakdown of Gaussian concentration occurs at longer times, which may be related to dynamical phase transitions.

From a numerical perspective, our algorithms should be compared in practice to existing approaches such as the closely related numerical linked-cluster expansion [53], which has also been used to approximate quantum dynamics [54–57]. Other related methods are cluster expansions with tensor-network representations [58,59] and schemes based on operator-basis expansions [60–62]. While our approach can be adapted to evolution under local Lindbladians by vectorizing the density operator, there is no obvious way in which noise improves convergence of the cluster expansion. It is unclear whether this happens for particular noise models, which could have significant implications on the classical simulation of noisy quantum circuits [63–65] and on the assessment of quantum advantage of noisy intermediate-scale quantum (NISQ) simulators [66].

We show that cluster expansion is useful not only for the study of systems in thermal equilibrium but also for dynamical problems. Cluster expansion enables us to establish the classical approximability of continuous dynamics for short times, similar to previous results for quantum circuits [67]. It complements other locality-based methods such as those derived from Lieb-Robinson bounds, which have led to important results in both dynamical and equilibrium systems [68]. We hope that our work stimulates further research into these techniques and into how they can help us understand other aspects of quantum many-body problems.

ACKNOWLEDGMENTS

APPENDIX A: COMPUTATION OF NESTED COMMUTATORS

We show in this appendix that the nested commutator in Eq. (9) can be numerically evaluated in time $\exp(\mathcal{O}(M))$. We start by expanding the commutator as

$$\sum_{\sigma \in S_m} \left[h_{W_{\sigma(1)}}, \cdots \left[h_{W_{\sigma(2)}}, \cdots \left[h_{W_{\sigma(m)}}, A \right] \right] \right]$$

$$= \sum_{I \subseteq [m]} (-1)^{m-l} \binom{m}{l} \left(\sum_{\sigma \in S_l} h_{W_{I_{\sigma(1)}}} \cdots h_{W_{I_{\sigma(l)}}} \right) A$$

$$\times \left(\sum_{\sigma \in S_{m-l}} h_{W_{J_{\sigma(1)}}} \cdots h_{W_{J_{\sigma(m-l)}}} \right), \qquad (A1)$$

where the sum over I runs over all subsets of $[m] = \{1, 2, \ldots, m\}$, $l = |I|$, and $J = [m] \setminus I$. The elements of I and J are labeled by $I_1, I_2, \ldots, I_l$ and $J_1, J_2, \ldots, J_{m-l}$ with some arbitrary ordering. Following Ref. [11], we use an inclusion-exclusion argument to rewrite each sum over permutations using the identity

$$\sum_{\sigma \in S_l} h_{W_{\sigma(1)}} \cdots h_{W_{\sigma(l)}} = \sum_{I \subseteq [l]} (-1)^{l-|I|} \left(\sum_{i \in I} h_{W_i} \right)^l. \quad (A2)$$

Hence,

$$\sum_{\sigma \in S_m} \left[h_{W_{\sigma(1)}}, \cdots \left[h_{W_{\sigma(2)}}, \cdots \left[h_{W_{\sigma(m)}}, A \right] \right] \right]$$

$$= \sum_{I \subseteq [m]} (-1)^l \binom{m}{l} \sum_{I' \subseteq I} \sum_{J' \subseteq J} (-1)^{|I'|+|J'|}$$

$$\times \left(\sum_{i \in I'} h_{W_i} \right)^l A \left(\sum_{i \in J'} h_{W_i} \right)^{m-l}. \qquad (A3)$$

We highlight that the sum over the subsets of $[m]$ involves 2^m terms as opposed to the $m!$ terms of the sum over the permutations in S_m. For a k-local Hamiltonian, the sums over h_{W_i} result in an operator that has support on at most km spins. The writing down of these operators on the relevant subspace can be achieved in time $\text{poly}(d^{km}) = \exp(\mathcal{O}(m))$. We can also raise them to the kth power with similar computational effort by diagonalizing the operators and powering the eigenvalues. Multiplying the resulting operators by A and ρ and taking the trace again incurs a computational cost with the same asymptotic dependence

on m. Finally, we need to perform the sums over I, I', and J'. There is a total of 4^m terms in these sums such that the overall computational effort indeed scales as $\exp(\mathcal{O}(m))$.

APPENDIX B: LOSCHMIDT ECHO

1. Illustrative example

In this appendix, we describe the lowest-order terms of the cluster expansion for the Loschmidt echo of a 2-local Hamiltonian. We emphasize that this example serves merely an illustrative purpose. Our results hold for the much broader class of Hamiltonians defined in Sec. II.

The spins in this example are arranged on a square lattice as indicated by the black circles in Fig. 4(a). The Hamiltonian is assumed to be a sum of terms that act on nearest-neighbor pairs. We note that the Hamiltonian may also include single-spin terms, as these can be absorbed into the 2-local interaction terms. Each interaction term of the Hamiltonian is represented in Fig. 4(a) by a light-blue diamond placed on the edges of the square lattice. To construct the interaction graph G, we connect two diamonds if their associated edges share a spin [dashed lines in Fig. 4(a)].

Connected clusters correspond to connected subgraphs of the interaction graphs. All possible connected clusters up to translations and rotations of sizes 1, 2, and 3 are shown in Figs. 4(b) and 4(c).

To gain intuition for what kind of terms appear in the cluster expansion, we consider the Taylor-series expansion of $\log\langle e^{-iHt}\rangle$ around $t = 0$. To third order,

$$\log\langle e^{-iHt}\rangle = -i\langle H\rangle t - \frac{1}{2}\left(\langle H^2\rangle - \langle H\rangle^2\right)t^2$$
$$+ \frac{i}{6}\left(\langle H^3\rangle - 3\langle H\rangle\langle H^2\rangle + \langle H\rangle^3\right)t^3 + O(t^4).$$
$$\text{(B1)}$$

The terms in the series take the form of cumulants. By assumption, the expectation value is with respect to a product state. This leads to a great number of cancellations when substituting in $H = \sum_X \lambda_X h_X$ because expectation

values of nonoverlapping terms factorize. For example, the term at second order simplifies to

$$\langle H^2\rangle - \langle H\rangle^2 = \sum_X \lambda_X^2\langle h_X^2\rangle + \sum_{\substack{X \neq Y \\ X \cap Y \neq \emptyset}} \lambda_X\lambda_Y\langle h_X h_y\rangle. \quad \text{(B2)}$$

We recognize the two sums as the two distinct types of connected clusters in Fig. 4(c), whereas all disconnected clusters cancel. Our formalism of the cluster expansion enables efficient book-keeping of these cancellations at arbitrary order.

2. Proof of Lemma 8

Formally, the cluster expansion of the Loschmidt echo is given by

$$L(t) = 1 + \sum_{m \geq 1}\sum_{\mathbf{W} \in \mathcal{C}_m}\frac{\lambda^{\mathbf{W}}}{\mathbf{W}!}\mathcal{D}_{\mathbf{W}}L(t). \quad \text{(B3)}$$

Note that the sum at this point includes both connected and disconnected clusters. The cluster derivative of $L(t)$ can be written as

$$\mathcal{D}_{\mathbf{W}}L(t) = (-it)^{|\mathbf{W}|}\langle h^{\mathbf{W}}\rangle_s, \quad \text{(B4)}$$

where we remind the reader that the symmetrized expectation value is defined by

$$\langle h^{\mathbf{W}}\rangle_s = \frac{1}{m!}\sum_{\sigma \in S_m}\text{tr}\left(h_{W_{\sigma(1)}}h_{W_{\sigma(2)}}\cdots h_{W_{\sigma(m)}}\rho\right), \quad \text{(B5)}$$

in which $m = |\mathbf{W}|$. The symmetrized expectation value has the property that it factorizes when $\mathbf{W}$ is disconnected. Denote by $P_{c,\max}(\mathbf{W}) \in \mathcal{P}_c(\mathbf{W})$ the partition of $\mathbf{W}$ into its
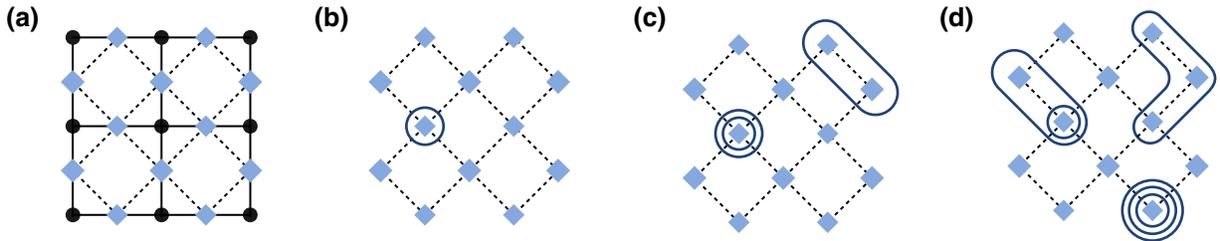
FIG. 4. (a) The spins (black circles) on a square lattice. The nearest-neighbor interaction is indicated by the light-blue diamonds. The dashed lines correspond to edges of the interaction graph. (b)–(d) Illustrations of all the connected clusters of size 1, 2, and 3, respectively, up to rotations and translations. The number of times an interaction term appears in a cluster is determined by how many dark-blue lines enclose it.

maximal connected components. Then,

$$\langle h^{\mathbf{W}} \rangle_s = \prod_{\mathbf{V} \in P_{c,\max}(\mathbf{W})} \langle h^{\mathbf{V}} \rangle_s \qquad (B6)$$

and

$$\frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} L(t) = \prod_{\mathbf{V} \in P_{c,\max}(\mathbf{W})} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(t). \qquad (B7)$$

The cluster expansion of the Loschmidt echo becomes

$$L(t) = 1 + \sum_{m \geq 1} \sum_{\mathbf{W} \in \mathcal{C}_m} \prod_{\mathbf{V} \in P_{c,\max}(\mathbf{W})} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(t). \qquad (B8)$$

At this point, we take the logarithm, which we again express in terms of its formal Taylor series,

$$\log(1+z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n. \qquad (B9)$$

Combining Eqs. (B8) and (B9) yields

$$\log L(t) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum_{\substack{m_1,\dots,m_n \geq 1}} \sum_{\substack{\mathbf{W}_1 \in \mathcal{C}_{m_1} \\ \cdots \\ \mathbf{W}_n \in \mathcal{C}_{m_n}}}$$
$$\times \left(\prod_{\mathbf{V}_1 \in P_{c,\max}(\mathbf{W}_1)} \frac{\lambda^{\mathbf{V}_1}}{\mathbf{V}_1!} \mathcal{D}_{\mathbf{V}_1} L(t) \right)$$
$$\cdots \left(\prod_{\mathbf{V}_n \in P_{c,\max}(\mathbf{W}_n)} \frac{\lambda^{\mathbf{V}_n}}{\mathbf{V}_n!} \mathcal{D}_{\mathbf{V}_n} L(t) \right). \qquad (B10)$$

We can rearrange the sums to first sum over all clusters $\mathbf{W}$ before considering decompositions of $\mathbf{W}$ into connected clusters $\mathbf{V}$.

$$\log L(t) = \sum_{m \geq 1} \sum_{\mathbf{W} \in \mathcal{G}_m} \sum_{P \in \mathcal{P}_c(\mathbf{W})} C(P) \prod_{\mathbf{V} \in P} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(t). \quad (B11)$$

Lemma 7 allows us to impose that $\mathbf{W}$ be connected. The coefficient $C(P)$ can be determined by considering the different ways in which the partition P can arise from the clusters $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n$ in Eq. (B10), such that $P = \bigcup_{i=1}^{n} P_{c,\max}(\mathbf{W}_i)$. Different elements of P can belong to the same "parent" cluster $\mathbf{W}_i$ if they do not overlap. It is hence possible to construct assignments of all $\mathbf{V} \in P$ to parent clusters $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n$ from a proper coloring of the partition graph $\tilde{G}_P$ with exactly n colors. The vertices colored

with the first color form $\mathbf{W}_1$, the second color gives $\mathbf{W}_2$, and so on. From this argument, we find that

$$C(P) = \frac{1}{P!} \sum_{n=1}^{|P|} \frac{(-1)^{n-1}}{n} \chi^*_{\tilde{G}_P}(n), \qquad (B12)$$

where $\chi^*_{\tilde{G}_P}(n)$ is the number of proper colorings of $\tilde{G}_P$ with exactly n colors. The combinatorial factor $P!$ removes overcounting that occurs when P contains clusters with multiplicity greater than 1 because permuting the colors of repeated clusters has no effect on $\mathbf{W}_1, \mathbf{W}_2, \dots \mathbf{W}_n$.

To complete the proof of Lemma 8, we make use of the following combinatorial property of graphs, which we prove in Appendix B 3.

Lemma 17. *Given a connected graph $G = (V, E)$, we denote by $\chi^*_G(n)$ the number of proper colorings of G with exactly n colors. Let $T_G(x, y)$ be the Tutte polynomial of G. Then,*

$$\sum_{n=1}^{|V|} \frac{(-1)^{n-1}}{n} \chi^*_G(n) = (-1)^{|V|-1} T_G(1, 0). \qquad (B13)$$

By applying this lemma to $C(P)$, we obtain

$$\log L(t) = \sum_{m \geq 1} \sum_{\mathbf{W} \in \mathcal{G}_m} \sum_{P \in \mathcal{P}_c(\mathbf{W})} \frac{(-1)^{|P|-1}}{P!} T_{\tilde{G}_P}(1, 0)$$
$$\times \prod_{\mathbf{V} \in P} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(t). \qquad (B14)$$

Finally, taking the cluster derivative of Eq. (B14) yields

$$\mathcal{D}_{\mathbf{W}} \log L(t) = \mathbf{W}! \sum_{P \in \mathcal{P}_c(\mathbf{W})} \frac{(-1)^{|P|-1}}{P!} T_{\tilde{G}_P}(1, 0)$$
$$\times \prod_{\mathbf{V} \in P} \frac{1}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(t), \qquad (B15)$$

which, using Eq. (B4), can be readily brought into the form of the expression in Lemma 8.

3. Proof of Lemma 17

To prove Lemma 17, we introduce three more short lemmas. The first one relates the number of colorings that use exactly k colors to the chromatic polynomial.

Lemma 18. *Given a graph G, let $\chi^*_G(k)$ denote the number of proper colorings of G that use exactly k colors. Moreover, let $\chi_G(k)$ be the chromatic polynomial, that is, the number of proper colorings with up to k colors. Then,*

$$\chi^*_G(n) = \sum_{k=1}^{n} (-1)^{n-k} \binom{n}{k} \chi_G(k). \qquad (B16)$$

Proof. This lemma follows from a standard inclusion-exclusion argument. Alternatively, we can prove the statement by direct calculation as follows.

The chromatic polynomial $\chi_G(k)$ can be computed by picking $j \leq k$ colors and adding the contributions from $\chi_G^*(j)$:

$$\chi_G(k) = \sum_{j=1}^{k} \binom{k}{j} \chi_G^*(j). \tag{B17}$$

We substitute this expression into the right-hand side of Eq. (B16) and exchange the order of the sums:

$$\sum_{k=1}^{n}(-1)^{n-k}\binom{n}{k}\chi_G(k) = \sum_{k=1}^{n}\sum_{j=1}^{k}(-1)^{n-k}\binom{n}{k}\binom{k}{j}\chi_G^*(j) = \sum_{j=1}^{n}\chi_G^*(j)\sum_{k=j}^{n}(-1)^{n-k}\binom{n}{k}\binom{k}{j}. \tag{B18}$$

When $j = n$, the last sum evaluates to 1. For $j < n$, we find instead that

$$\sum_{k=j}^{n}(-1)^{n-k}\binom{n}{k}\binom{k}{j} = \binom{n}{j}\sum_{k=j}^{n}(-1)^{n-k}\binom{n-j}{k-j} = \binom{n}{j}\sum_{k=0}^{n-j}(-1)^{n-j-k}\binom{n-j}{k} = 0. \tag{B19}$$

Thus,

$$\sum_{k=1}^{n}(-1)^{n-k}\binom{n}{k}\chi_G(k) = \chi_G^*(n), \tag{B20}$$

as claimed in the lemma. ∎

The second lemma connects the chromatic polynomial to the Tutte polynomial. A proof of this statement can be found in, e.g., Ref. [69].

Lemma 19. *Given a graph $G = (V, E)$ with c connected components, the chromatic polynomial $\chi_G(k)$ is related to the Tutte polynomial $T_G(x, y)$ by*

$$\chi_G(k) = (-1)^{|V|-c}k^c T_G(1-k, 0). \tag{B21}$$

The third lemma is a simple identity involving sums over binomial coefficients and powers of integers.

Lemma 20. *The following identity holds for any integers n and k satisfying $n > k \geq 0$.*

$$\sum_{j=1}^{n}(-1)^{n-j}\binom{n}{j}j^k = (-1)^{n-1}\delta_{k,0}. \tag{B22}$$

Proof. For $k = 0$,

$$\sum_{j=1}^{n}(-1)^{n-j}\binom{n}{j} = (-1)^{n-1} + \sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j} = (-1)^{n-1}. \tag{B23}$$

For $n > k > 0$, we observe that

$$\sum_{j=1}^{n}(-1)^{n-j}\binom{n}{j}j^k x^j = \left(x\frac{d}{dx}\right)^k (x-1)^n. \tag{B24}$$

The lemma follows from the fact that the right-hand side vanishes at $x = 1$ for all $n > k > 0$. ∎

Proof of Lemma 17. We apply the above three lemmas in order. From Lemma 18, we have

$$\sum_{n=1}^{|V|}\frac{(-1)^{n-1}}{n}\chi_G^*(n) = \sum_{n=1}^{|V|}\frac{(-1)^{n-1}}{n}\sum_{k=1}^{n}(-1)^{n-k}\binom{n}{k}\chi_G(k). \tag{B25}$$

We switch the order of the sums to obtain

$$\sum_{n=1}^{|V|}\frac{(-1)^{n-1}}{n}\chi_G^*(n) = \sum_{k=1}^{|V|}\sum_{n=k}^{|V|}\frac{(-1)^{k-1}}{n}\binom{n}{k}\chi_G(k)$$

$$= \sum_{k=1}^{|V|}\frac{(-1)^{k-1}}{k}\chi_G(k)\sum_{n=k}^{|V|}\binom{n-1}{k-1}. \tag{B26}$$

By the hockey-stick identity, the last sum evaluates to $\binom{|V|}{k}$. Combined with Lemma 19, setting $c = 1$ since G is connected by assumption, this yields

$$\sum_{n=1}^{|V|}\frac{(-1)^{n-1}}{n}\chi_G^*(n) = \sum_{k=1}^{|V|}\binom{|V|}{k}(-1)^{|V|-k}T_G(1-k, 0). \tag{B27}$$

The Tutte polynomial $T_G(x, 0)$ is a polynomial in x of degree at most $|V| - 1$. Therefore, $T_G(1 - k, 0)$ is a polynomial in k of the same maximum degree, which

allows us to write

$$T_G(1 - k, 0) = \sum_{n=0}^{|V|-1} a_n k^n \qquad (B28)$$

for some coefficients a_n. By substituting into Eq. (B27), we obtain

$$\sum_{n=1}^{|V|} \frac{(-1)^{n-1}}{n} \chi_G^*(n) = \sum_{n=0}^{|V|-1} a_n \sum_{k=1}^{|V|} \binom{|V|}{k} (-1)^{|V|-k} k^n. \qquad (B29)$$

Lemma 20 allows us to simplify the expression to

$$\sum_{n=1}^{|V|} \frac{(-1)^{n-1}}{n} \chi_G^*(n) = (-1)^{|V|-1} a_0. \qquad (B30)$$

This is the desired expression, since $a_0 = T_G(1, 0)$. ∎

4. Proof of Proposition 9

We start from the expression for the cluster derivative in Lemma 8:

$$\mathcal{D}_{\mathbf{W}} \log L(t) = (-it)^m \sum_{P \in \mathcal{P}_c(\mathbf{W})} (-1)^{|P|-1} N_P(\mathbf{W}) T_{\tilde{G}_P}(1, 0)$$

$$\prod_{\mathbf{V} \in P} \langle h^{\mathbf{V}} \rangle_s.$$

Recall the symmetric expectation value $\langle h^{\mathbf{V}} \rangle_s = 1/|\mathbf{V}|!$ $\sum_{\sigma \in S_{|\mathbf{V}|}} \operatorname{tr} \left(h_{V_{\sigma(1)}} h_{V_{\sigma(2)}} \cdots h_{V_{\sigma(|\mathbf{V}|)}} \rho \right)$ and the combinatorial factor $N_P(\mathbf{W}) = \mathbf{W}! / \left(P! \prod_{\mathbf{V} \in P} \mathbf{V}! \right)$. We rewrite the sum over cluster partitions of $\mathbf{W}$ as a sum of graph partitions of the cluster graph $G_{\mathbf{W}}$. Here, a graph partition refers to a partition of the vertices. We only consider partitions into connected subgraphs, meaning that the subsets of vertices in the partition induce connected subgraphs on the original graph. For every graph partition of $G_{\mathbf{W}}$ into connected subgraphs, there exists exactly one corresponding cluster partition $P \in \mathcal{P}_c(\mathbf{W})$. On the other hand, for every cluster partition, there are exactly $N_P(\mathbf{W}) = \mathbf{W}! / P! \prod_{\mathbf{V} \in P} \mathbf{V}!$ equivalent graph partitions. This combinatorial factor arises because repeated subsystems are indistinguishable at the level of the cluster but give rise to distinct vertices in the cluster graph. With this,

$$\mathcal{D}_{\mathbf{W}} \log L(t) = (-it)^m \sum_{P' \in \mathcal{P}_c(G_{\mathbf{W}})} (-1)^{|P'|-1}$$

$$T_{\tilde{G}_P}(1, 0) \prod_{\mathbf{V} \in P} \langle h^{\mathbf{V}} \rangle_s, \qquad (B31)$$

where, in a slight abuse of notation, $\mathcal{P}_c(G_{\mathbf{W}})$ is the set of partitions of the cluster graph $G_{\mathbf{W}}$ into connected components and $P \in \mathcal{P}_c(\mathbf{W})$ is the cluster partition corresponding

to P'. By observing that $|\langle h^{\mathbf{V}} \rangle_s| \leq 1$, we obtain

$$|\mathcal{D}_{\mathbf{W}} \log L(t)| \leq |t|^m \sum_{P' \in \mathcal{P}_c(G_{\mathbf{W}})} T_{\tilde{G}_P}(1, 0) \leq |t|^m$$

$$\sum_{P' \in \mathcal{P}_c(G_{\mathbf{W}})} T_{\tilde{G}_P}(1, 1), \qquad (B32)$$

where the last inequality follows from the fact that the Tutte polynomial $T_{\tilde{G}_P}(x, y)$ has positive coefficients. We bound this sum in two steps, starting with the following lemma.

Lemma 21. *Given a connected cluster $\mathbf{W} \in \mathcal{G}_m$,*

$$\sum_{P' \in \mathcal{P}_c(G_{\mathbf{W}})} T_{\tilde{G}_P}(1, 1) \leq 2^m T_{G_{\mathbf{W}}}(1, 1). \qquad (B33)$$

Proof. $T_{\tilde{G}_{P'}}(1, 1)$ counts the number of spanning trees of $\tilde{G}_P$, so that

$$\sum_{P' \in \mathcal{P}_c(G_{\mathbf{W}})} T_{\tilde{G}_P}(1, 1) = \sum_{P' \in \mathcal{P}'_c(G_{\mathbf{W}})} \sum_{\substack{\text{spanning} \\ \text{trees of } \tilde{G}_P}} 1. \qquad (B34)$$

Given a spanning tree of the cluster graph $G_{\mathbf{W}}$, consider a bicoloring of the edges into blue and red and delete the red ones. This separates the edges into disconnected components, each of which induces a connected subgraph on $G_{\mathbf{W}}$. These subgraphs define a partition $P' \in \mathcal{P}_c(G_{\mathbf{W}})$, with its corresponding partition graph $\tilde{G}_P$. Moreover, the deleted red edges can be identified with a spanning tree of $\tilde{G}_P$. Hence, any bicoloring of a spanning tree of $G_{\mathbf{W}}$ describes a term in the double sum on the right-hand side of Eq. (B34). Conversely, for every term in the sum, we can find at least one bicolored spanning tree. This procedure is illustrated in Fig. 5. It follows that the sum is bounded from above by the number of bicolored trees of $G_{\mathbf{W}}$. The number of edges of each spanning tree is $|\mathbf{W}| - 1 = m - 1$, so that there are always 2^{m-1} distinct bicolorings for every tree. The total number of spanning trees is $T_{G_{\mathbf{W}}}(1, 1)$, which completes the proof. ∎

We next bound the number of spanning trees of $T_{G_{\mathbf{W}}}(1, 1)$ given the connectivity of the cluster graph.

Lemma 22. *Consider a set of supports S such that the maximum degree of the associated interaction graph G is $\mathfrak{d}$. Given a connected cluster $\mathbf{W} \in \mathcal{G}_m$,*

$$\frac{1}{\mathbf{W}!} T_{G_{\mathbf{W}}}(1, 1) \leq [e(\mathfrak{d} + 1)]^{m+1}. \qquad (B35)$$

Proof. Let us fix $W_1 \in \mathbf{W}$ as the root of a spanning tree. Any spanning tree can be constructed by picking for each
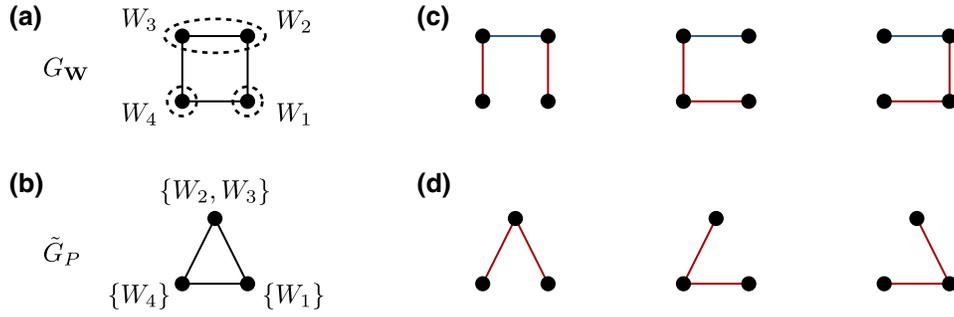
FIG. 5. (a) A cluster graph $G_{\mathbf{W}}$ with a particular graph partition indicated by the dashed outlines. (b) The corresponding partition graph. (c) The three bicolored spanning trees that we identify with this graph partition. (d) The red edges define a spanning tree on the partition graph.

vertex $W_2, W_3, \ldots W_m$ one of the edges incident on it. The number of choices is $\gamma(W_2)\gamma(W_3)\cdots\gamma(W_m)$, where $\gamma(W_k)$ is the degree of the vertex associated with W_k in $G_{\mathbf{W}}$. It follows that

$$T_{G_{\mathbf{W}}}(1,1) \leq \gamma(W_1)\gamma(W_2)\cdots\gamma(W_m). \qquad \text{(B36)}$$

This product of degrees can be bounded as in the proof of Ref. [22, Proposition 3.8]. The degree of a vertex associated with W_k can be written as

$$\gamma(W_k) = \mu_{\mathbf{W}}(W_k) - 1 + \sum_{X \in \mathcal{N}(W_k)} \mu_{\mathbf{W}}(X), \qquad \text{(B37)}$$

where $\mathcal{N}(W_k)$ is the set of neighbors of W_k in the interaction graph G. We sum the degrees over all distinct subsystems that appear in $\mathbf{W}$:

$$\sum_{X \in S \,:\, \mu_{\mathbf{W}}(X)>0} \gamma(X) = \sum_{X \in S \,:\, \mu_{\mathbf{W}}(X)>0}$$

$$\times \left(\mu_{\mathbf{W}}(X) - 1 + \sum_{Y \in \mathcal{N}(X)} \mu_{\mathbf{W}}(Y)\right) \leq m - 1 + \mathfrak{d}m. \qquad \text{(B38)}$$

The double sum is bounded by $\mathfrak{d}m$ because each $\mu_{\mathbf{W}}(Y)$ appears in it at most $\mathfrak{d}$ times. Finally, we bound

$$\frac{1}{\mathbf{W}!}\gamma(W_1)\gamma(W_2)\cdots\gamma(W_m)$$

$$= \prod_{X \in S \,:\, \mu_{\mathbf{W}}(X)>0} \frac{\left(\mu_{\mathbf{W}}(X) - 1 + \sum_{Y \in \mathcal{N}(X)} \mu_{\mathbf{W}}(Y)\right)^{\mu_{\mathbf{W}}(X)}}{\mu_{\mathbf{W}}(X)!} \qquad \text{(B39)}$$

$$\leq e^m \prod_{X \in S \,:\, \mu_{\mathbf{W}}(X)>0} \left(\frac{\mu_{\mathbf{W}}(X) - 1 + \sum_{Y \in \mathcal{N}(X)} \mu_{\mathbf{W}}(Y)}{\mu_{\mathbf{W}}(X)}\right)^{\mu_{\mathbf{W}}(X)} \qquad \text{(B40)}$$

$$\leq e^m \left(\frac{m - 1 + \mathfrak{d}m}{m}\right)^m \leq [e(\mathfrak{d}+1)]^m. \qquad \text{(B41)}$$

∎

The bound on the cluster derivative follows from the two lemmas and Eq. (B32).

APPENDIX C: GENERALIZED LOSCHMIDT ECHO

In this appendix, we prove Theorem 13. To this end, we first establish Eq. (44). We again start from the formal cluster expansion of the Loschmidt echo:

$$L(\{t_l\}) = 1 + \sum_{m \geq 1} \sum_{\mathbf{W} \in \mathcal{C}_m^K} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} L(\{t_l\}). \qquad \text{(C1)}$$

The cluster derivative takes the more complicated form

$$
\mathcal{D}_{\mathbf{W}} L(\{t_l\}) = \left[\prod_{l=1}^{K} \frac{(-it_l)^{m_l}}{m_l!}\right] \mathrm{tr}
$$
$$
\times \left[\left(\sum_{\sigma \in S_{m_1}} h^{(1)}_{W_{1,\sigma(1)}} \cdots h^{(1)}_{W_{1,\sigma(m_1)}}\right)\right.
$$
$$
\left.\cdots \left(\sum_{\sigma \in S_{m_K}} h^{(K)}_{W_{K,\sigma(1)}} \cdots h^{(K)}_{W_{K,\sigma(m_K)}}\right) \rho\right].
$$
(C2)

Here, $m_l = |\mathbf{W}_l|$ and $\mathbf{W}_l$ are the parts of $\mathbf{W}$ associated with the Hamiltonian $H^{(l)}$. Despite these complications, one can readily check that we still have

$$
\frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} L(\{t_l\}) = \prod_{\mathbf{V} \in P_{c,\max}(\mathbf{W})} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(\{t_l\}).
$$
(C3)

Here, the components are disconnected if their subsystems do not overlap, irrespective of the Hamiltonian label. The remaining arguments from Appendix B 2 carry over and we obtain

$$
\frac{1}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(\{t_l\}) = \sum_{P \in \mathcal{P}_c(\mathbf{W})} \frac{(-1)^{|P|-1}}{P!} T_{\tilde{G}_P}(1,0)
$$
$$
\times \prod_{\mathbf{V} \in P} \frac{1}{\mathbf{V}!} \mathcal{D}_{\mathbf{V}} L(\{t_l\}).
$$
(C4)

By using the fact that $|\mathcal{D}_{\mathbf{W}} L(\{t_l\})| \le \prod_{l=1}^{K} |t_l|^{m_l}$ and following the steps in Appendix B 4, we arrive at an upper bound for the cluster derivative. The only change is that the relevant interaction graph is G^K, which has maximum degree $K(\mathfrak{d}+1)-1$ given the maximum degree $\mathfrak{d}$ of G. For $\mathbf{W} \in \mathcal{G}_m^K$, this yields Eq. (44).

To prove the convergence statement in Theorem 13, we observe that

$$
\sum_{\mathbf{W} \in \mathcal{G}_m^K} \prod_{l=1}^{K} |t_l|^{m_l} \le \sum_{\mathbf{W} \in \mathcal{G}_m} \sum_{\substack{m_1,\ldots,m_K \ge 0 \\ m_1+\cdots+m_K=m}} \frac{m!}{m_1! \cdots m_K!} \prod_{l=1}^{K} |t_l|^{m_l}
$$
$$
= \left(\sum_{l=1}^{K} |t_l|\right)^m |\mathcal{G}_m| \le |S| \left(e\mathfrak{d} \sum_{l=1}^{K} |t_l|\right)^m,
$$
(C5)

where we use the bound on $|\mathcal{G}_m|$ from Lemma 1. Combining these results yields

$$
\left|\sum_{\mathbf{W} \in \mathcal{G}_m^K} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(t)\right| \le |S| \left[2e^2 K \mathfrak{d}(\mathfrak{d}+1) \sum_{l=1}^{K} |t_l|\right]^m.
$$
(C6)

The convergence of the cluster expansion for this generalized Loschmidt echo follows in an analogous fashion to Theorem 10.

To prove the bound on the computational cost in Theorem 13, we closely follow the proofs of Proposition 11 and Theorem 12, keeping in mind that the degree of the relevant interaction graph is $K(\mathfrak{d}+1)-1$. Throughout, we only keep the dependence on K explicit, while suppressing the dependence on $\mathfrak{d} = \mathcal{O}(1)$.

The computational cost of step (i) of the proof of Proposition 11 is modified to $\exp(\mathcal{O}(m \log K))$. Step (ii) remains unchanged, as the computational cost of evaluating the Tutte polynomial only depends on the number of vertices. In step (iii), we have to evaluate Eq. (C2) instead of the simpler symmetrized expectation value. Nevertheless, by rewriting the sums over permutations using Eq. (A2), this can still be carried out in time $\exp(\mathcal{O}(n))$. Hence, the cluster derivative of the generalized Loschmidt echo can be computed in time $\exp(\mathcal{O}(m \log K))$. By Lemma 1, the run time of the algorithm enumerating the clusters is $|S| \times \exp(\mathcal{O}(m \log K))$. We conclude that the truncated cluster expansion of the generalized Loschmidt echo that includes all clusters up to size M can be computed with a total run time $|S| \times \exp(\mathcal{O}(M \log K))$. Choosing $M > \log |S|/\left(1 - K\sum_l |t_l|/t_L^*\right) \varepsilon / \log t_L^*/K \sum_l |t_l|$ completes the proof.

APPENDIX D: PROOF OF CONCENTRATION BOUND

Let us assume that $\mathrm{tr}(\rho e^{itH^{(1)}} H^{(2)} e^{-itH^{(1)}}) = 0$ for simplicity. We write the cluster expansion of $\log L(\{t, \nu, -t\}) \equiv \log \mathrm{tr}(\rho e^{itH^{(1)}} e^{\nu H^{(2)}} e^{-itH^{(1)}})$ as

$$
f_L(\{t,\nu,-t\}) = \log \mathrm{tr}(\rho e^{itH^{(1)}} e^{\nu H^{(2)}} e^{-itH^{(1)}})
$$
$$
= \sum_{m \ge 1} \sum_{\mathbf{W} \in \mathcal{G}_m^3} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(\{t,\nu,-t\}), \quad \text{(D1)}
$$

where we have three Hamiltonians. However, note that since $f_L(\{t,0,-t\}) = 0$, only the clusters with at least two terms $h_X^{(2)}$ from $H^{(2)}$ contribute. This means that the smallest power of ν is 2. That is,

$$
f_L(\{t,\nu,-t\}) = \sum_{m \ge 2} \sum_{\mathbf{W} \in \mathcal{G}_m^3: \, m_2 \ge 2} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L(\{t,\nu,-t\}),
$$
(D2)

where m_2 is the number of subsystems in $\mathbf{W}$ associated with $H^{(2)}$. Now, given Eq. (44) and the bound on the number of clusters in Lemma 1, we have that

$$
|f_L(\{t,\nu,-t\})| \le \sum_{m \ge 2} \sum_{\mathbf{W} \in \mathcal{G}_m^3: \, m_2 \ge 2} |t|^{m-m_2} \nu^{m_2} (6e(\mathfrak{d}+1))^m
$$
(D3)

$$\leq |S| \sum_{m \geq 2} (6e(\mathfrak{d} + 1))^m (e\mathfrak{d})^m \left(\sum_{\substack{m_1, m_3 \geq 0, \ m_2 \geq 2 \\ m_1 + m_2 + m_3 = m}} \frac{m!}{m_1! m_2! m_3!} |t|^{m - m_2} \nu^{m_2} \right) \tag{D4}$$

$$\leq |S| \sum_{m \geq 2} (6e^2 \mathfrak{d}(\mathfrak{d} + 1))^m \nu^2 m^2 \left(\sum_{\substack{m_1, m_2, m_3 \geq 0 \\ m_1 + m_2 + m_3 = m - 2}} \frac{(m - 2)!}{m_1! m_2! m_3!} |t|^{m - m_2} \nu^{m_2} \right) \tag{D5}$$

$$= |S| \sum_{m \geq 2} (6e^2 \mathfrak{d}(\mathfrak{d} + 1))^m \nu^2 m^2 (2|t| + \nu)^{m-2}, \tag{D6}$$

where in the second line we write the sum over combinations with at least two terms from $H^{(2)}$ (i.e., the number of ways in which one can arrange m objects in three boxes, with at least two objects in one of them). In the third line, we bound this sum by m^2 times the number of ways of arranging $m - 2$ elements in three boxes. Finally, we use $\sum_{m \geq 2} m^2 a^{m-2} = 4 + a^2 - 3a/(1-a)^3$ and the definition of t_L^* in Theorem 10 to obtain

$$f_L(\{t, \nu, -t\}) \leq |S| \left(\frac{3\nu}{t_L^*}\right)^2 \frac{4 + (3(2|t| + \nu)/t_L^*)^2}{\left[1 - 3(2|t| + \nu)/t_L^*\right]^3}. \tag{D7}$$

Now take $|t| \leq t_L^*/7$, so that $1 - 3(2|t| + \nu)/t_L^* \geq 1/7 - 3\nu/t_L^*$. We obtain

$$\log \operatorname{tr}(\rho(t) e^{\nu H^{(2)}}) \leq 5|S| \frac{(3\nu/t_L^*)^2}{\left(1/7 - 3\nu/t_L^*\right)^3}. \tag{D8}$$

Choosing $\nu = \eta \delta (t_L^*)^2 / |S|$ with $\eta = 1/(90 \times 7^3)$ then leads to

$$e^{-\delta \nu} \operatorname{tr}\left[\rho(t) e^{\nu H^{(2)}}\right] \leq e^{-(\delta t_L^*)^2/(250)^2 |S|}, \tag{D9}$$

under the condition that $\delta \leq |S|/t_L^*$, which is trivially satisfied since $\delta \leq |S|$ and $t_L^* < 1$. Together with the inequality

$$\Pr[|x - \operatorname{tr}[\rho(t) H^{(2)}]| \geq \delta] \leq 2e^{-\delta \nu} \operatorname{tr}(\rho(t) e^{\nu H^{(2)}}), \tag{D10}$$

this proves the result.

APPENDIX E: PROOF OF COROLLARY 15

From Eq. (26) we have the series expansion

$$\frac{g_n(t)}{n} = \frac{1}{n} \left(\sum_{m=1}^{\infty} \sum_{\mathbf{W} \in \mathcal{G}_m} \frac{\lambda^{\mathbf{W}}}{\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L_n(t) \right). \tag{E1}$$

Theorem 10 shows that $g_n(t)$ is analytic for $t < t_L^*$ and any finite n, such that

$$\left| \sum_{\mathbf{W} \in \mathcal{G}_m} \frac{\lambda^{\mathbf{W}}}{n \times \mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L_n(t) \right| \leq \frac{e\mathfrak{d}|S|}{(1 + e(\mathfrak{d} - 1))n} (|t|/t_L^*)^m. \tag{E2}$$

For local Hamiltonians, $|S|/n \equiv C = \mathcal{O}(1)$ and thus

$$\sum_{m=m_0}^{\infty} \left(\frac{t}{t_L^*}\right)^m C e\mathfrak{d} < \infty. \tag{E3}$$

We apply Tannery's theorem to the sequence $\{\sum_{\mathbf{W} \in \mathcal{G}_m} \lambda^{\mathbf{W}} / n\mathbf{W}! \mathcal{D}_{\mathbf{W}} \log L_n(t)\}$. This implies that we can place the limit inside the sum as

$$G(t) = \sum_{m=0}^{\infty} \lim_{n \to \infty} \sum_{\mathbf{W} \in \mathcal{G}_m} \frac{\lambda^{\mathbf{W}}}{n\mathbf{W}!} \mathcal{D}_{\mathbf{W}} \log L_n(t). \tag{E4}$$

That $G(t)$ is analytic follows from the fact that $\lim_{n \to \infty} \sum_{\mathbf{W} \in \mathcal{G}_m} \lambda^{\mathbf{W}} / n\mathbf{W}! \mathcal{D}_{\mathbf{W}} \log L_n(t) \leq (t/t_L^*)^m C e\mathfrak{d}$.

[1] T. J. Osborne, Efficient Approximation of the Dynamics of One-Dimensional Quantum Spin Systems, Phys. Rev. Lett. **97**, 157202 (2006).

[2] T. Kuwahara, Á. M. Alhambra, and A. Anshu, Improved Thermal Area Law and Quasilinear Time Algorithm for Quantum Gibbs States, Phys. Rev. X **11**, 011047 (2021).

[3] N. P. Breuckmann and J. N. Eberhardt, Quantum Low-Density Parity-Check Codes, PRX Quantum **2**, 040101 (2021).

[4] D. Ruelle, *Statistical Mechanics* (Imperial College Press, London, 1999).

[5] S. Friedli and Y. Velenik, *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction* (Cambridge University Press, Cambridge, United Kingdom, 2017), 232.

[6] V. A. Malyshev, Cluster expansions in lattice models of statistical physics and the quantum theory of fields, Russ. Math. Surv. **35**, 1 (1980).

[7] Y.-M. Park, The cluster expansion for classical and quantum lattice systems, J. Stat. Phys. **27**, 553 (1982).

[8] R. Kotecký and D. Preiss, Cluster expansion for abstract polymer models, Commun. Math. Phys. **103**, 491 (1986).

[9] R. L. Dobrushin and S. B. Shlosman, Completely analytical interactions: Constructive description, J. Stat. Phys. **46**, 983 (1987).

[10] R. L. Dobrushin, in *Topics in Statistical and Theoretical Physics*, Vol. 177 (American Mathematical Society, Providence, Rhode Island, 1996), p. 59.

[11] R. L. Mann and T. Helmuth, Efficient algorithms for approximating quantum partition functions, J. Math. Phys. **62**, 022201 (2021).

[12] A. Harrow, S. Mehraban, and M. Soleimanifar, in *STOC 2020: Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2020), p. 378.

[13] D. Ueltschi, Cluster expansions and correlation functions, Mosc. Math. J. **4**, 511 (2004).

[14] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert, Locality of Temperature, Phys. Rev. X **4**, 031019 (2014).

[15] J. Fröhlich and D. Ueltschi, Some properties of correlations of quantum lattice systems in thermal equilibrium, J. Math. Phys. **56**, 053302 (2015).

[16] K. Netočný and F. Redig, Large deviations for quantum spin systems, J. Stat. Phys. **117**, 521 (2004).

[17] T. Kuwahara and K. Saito, Gaussian concentration bound and ensemble equivalence in generic quantum many-body systems including long-range interaction, Ann. Phys. **421**, 168278 (2020).

[18] D. Janzing and P. Wocjan, Ergodic quantum computing, Quantum Inf. Process. **4**, 129 (2005).

[19] A. Galanis, L. Ann Goldberg, and A. Herrera-Poyatos, The complexity of approximating the complex-valued Ising model on bounded degree graphs (2021), arXiv preprint, ArXiv:2105.00287.

[20] A. Galanis, L. Ann Goldberg, and A. Herrera-Poyatos, The complexity of approximating the complex-valued Potts model, Comput. Complex. **31**, 2 (2022).

[21] G. De las Cuevas, W Dür, M Van den Nest, and M. A. Martin-Delgado, Quantum algorithms for classical lattice models, New J. Phys. **13**, 093021 (2011).

[22] J. Haah, R. Kothari, and E. Tang, Optimal learning of quantum Hamiltonians from high-temperature Gibbs states (2021), arXiv preprint, ArXiv:2108.04842.

[23] M. Heyl, Dynamical quantum phase transitions: A review, Rep. Prog. Phys. **81**, 054001 (2018).

[24] S. Deffner and S. Campbell, Quantum speed limits: From Heisenberg's uncertainty principle to optimal quantum control, J. Phys. A: Math. Theor. **50**, 453001 (2017).

[25] L. Mandelstam and I. E. Tamm, in *Selected Papers*, edited by Boris M. Bolotovskii, Victor Ya. Frenkel, and Rudolf Peierls (Springer-Verlag, Berlin, 1991), p. 115.

[26] N. Margolus and L. B. Levitin, The maximum speed of dynamical evolution, Phys. D **120**, 188 (1998).

[27] T. Kuwahara, Connecting the probability distributions of different operators and generalization of the Chernoff-Hoeffding inequality, J. Stat. Mech.: Theory Exp. **2016**, 113103 (2016).

[28] A. Anshu, Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems, New J. Phys. **18**, 083011 (2016).

[29] A. Anshu and T. Metger, Concentration bounds for quantum states and limitations on the QAOA from polynomial approximations (2022), arXiv preprint, ArXiv:2209.02715.

[30] S. Bravyi, A. Chowdhury, D. Gosset, and P. Wocjan, Quantum Hamiltonian complexity in thermal equilibrium, Nat. Phys. (2022).

[31] A. Barvinok, *Combinatorics and Complexity of Partition Functions, Algorithms and Combinatorics* (Springer International Publishing, Cham, Switzerland, 2016).

[32] L. N. Trefethen, Quantifying the ill-conditioning of analytic continuation, BIT Numer. Math. **60**, 901 (2020).

[33] G. Bouch, Complex-time singularity and locality estimates for quantum lattice systems, J. Math. Phys. **56**, 123303 (2015).

[34] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, Commun. Math. Phys. **28**, 251 (1972).

[35] T. Gorin, T. Prosen, T. H. Seligman, and M. Žnidarič, Dynamics of Loschmidt echoes and fidelity decay, Phys. Rep. **435**, 33 (2006).

[36] J. Emerson, S. Lloyd, D. Poulin, and D. Cory, Estimation of the local density of states on a quantum computer, Phys. Rev. A **69**, 050305(R) (2004).

[37] S. Lu, and M. Carmen Bañuls, and J. I. Cirac, Algorithms for Quantum Simulation at Finite Energies, PRX Quantum **2**, 020321 (2021).

[38] N. Biggs, *Algebraic Graph Theory* (Cambridge University Press, Cambridge, United Kingdom, 1993), 2nd ed.

[39] T. Helmuth, W. Perkins, and G. Regts, Algorithmic Pirogov-Sinai theory, Probab. Theory Relat. Fields **176**, 851 (2020).

[40] T. Kuwahara, K. Kato, and F. G. S. L. Brandão, Clustering of Conditional Mutual Information for Quantum Gibbs States above a Threshold Temperature, Phys. Rev. Lett. **124**, 220601 (2020).

[41] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto, in *2008 49th Annual IEEE Symposium on Foundations of Computer Science* (IEEE, Philadelphia, Pennsylvania, 2008), p. 677.

[42] H. Wilming, T. R. de Oliveira, A. J. Short, and J. Eisert, Equilibration times in closed quantum many-body systems, Thermodyn. Quantum Regime: Fundam. Aspects New Directions (2019).

[43] T. Kuwahara and K. Saito, Eigenstate Thermalization from the Clustering Property of Correlation, Phys. Rev. Lett. **124**, 200604 (2020).

[44] G. De Palma, M. Marvian, C. Rouzé, and D. Stilck França, Limitations of Variational Quantum Algorithms: A Quantum Optimal Transport Approach, PRX Quantum **4**, 010309 (2023).

[45] G. De Palma and C. Rouzé, Quantum concentration inequalities, Ann. Henri Poincaré **23**, 3391 (2022).

[46] M. Heyl, Scaling and Universality at Dynamical Quantum Phase Transitions, Phys. Rev. Lett. **115**, 140602 (2015).

[47] L. Piroli, B. Pozsgay, and E. Vernier, Non-analytic behavior of the Loschmidt echo in XXZ spin chains: Exact results, Nucl. Phys. B **933**, 454 (2018).

[48] L. Campos Venuti and P. Zanardi, Unitary equilibrations: Probability distribution of the Loschmidt echo, Phys. Rev. A **81**, 022113 (2010).

[49] Á. M. Alhambra, A. Anshu, and H. Wilming, Revivals imply quantum many-body scars, Phys. Rev. B **101**, 205107 (2020).

[50] G. De Palma, M. Marvian, C. Rouzé, and D. Stilck França, Limitations of variational quantum algorithms: A quantum optimal transport approach (2022), arXiv preprint, ArXiv:2204.03455.

[51] R. Fernández and A. Procacci, Cluster expansion for abstract polymer models: New bounds from an old approach, Commun. Math. Phys. **274**, 123 (2007).

[52] A. H. Moosavian, S. S. Kahani, and S. Beigi, Limits of short-time evolution of local Hamiltonians, Quantum **6**, 744 (2022).

[53] M. Rigol, T. Bryant, and R. R. P. Singh, Numerical Linked-Cluster Approach to Quantum Lattice Models, Phys. Rev. Lett. **97**, 187202 (2006).

[54] M. Rigol, Quantum Quenches in the Thermodynamic Limit, Phys. Rev. Lett. **112**, 170601 (2014).

[55] I. G. White, B. Sundar, and K. R. A. Hazzard, Quantum dynamics from a numerical linked cluster expansion (2017), arXiv preprint, ArXiv:1710.07696.

[56] J. Gan and K. R. A. Hazzard, Numerical linked cluster expansions for inhomogeneous systems, Phys. Rev. A **102**, 013318 (2020).

[57] J. Richter, T. Heitmann, and R. Steinigeweg, Quantum quench dynamics in the transverse-field Ising model: A numerical expansion in linked rectangular clusters, SciPost Phys. **9**, 031 (2020).

[58] A. Molnár, N. Schuch, F. Verstraete, and J. I. Cirac, Approximating Gibbs states of local Hamiltonians efficiently with PEPS, Phys. Rev. B **91**, 045138 (2015).

[59] B. Vanhecke, D. Devoogdt, F. Verstraete, and L. Vanderstraeten, Simulating thermal density operators with cluster expansions and tensor networks (2021), arXiv preprint, ArXiv:2112.01507.

[60] C. David White, M. Zaletel, R. S. K. Mong, and G. Refael, Quantum dynamics of thermalizing systems, Phys. Rev. B **97**, 035127 (2018).

[61] T. K. Kvorning, L. Herviou, and J. H. Bardarson, Time-evolution of local information: Thermalization dynamics of local observables, SciPost Phys. **13**, 080 (2022).

[62] C. von Keyserlingk, F. Pollmann, and T. Rakovszky, Operator backflow and the classical simulation of quantum transport, Phys. Rev. B **105**, 245101 (2022).

[63] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation (1996), arXiv preprint, ArXiv:quant-ph/9611028.

[64] D. Stilck França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nat. Phys. **17**, 1221 (2021).

[65] G. González-García, R. Trivedi, and J. I. Cirac, Error propagation in NISQ devices for solving classical optimization problems (2022), arXiv preprint, ArXiv:2203.15632.

[66] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, Nature **607**, 667 (2022).

[67] S. Bravyi, D. Gosset, and R. Movassagh, Classical algorithms for quantum mean values, Nat. Phys. **17**, 337 (2021).

[68] M. B. Hastings, Locality in quantum systems (2010), arXiv preprint, ArXiv:1008.5137.

[69] T. Brylawski and J. Oxley, in *Matroid Applications* (Cambridge University Press, Cambridge, United Kingdom, 1992), p. 123.