

Circuit-Based Characterization of Finite-Temperature Quantum Phases and Self-Correcting Quantum Memory

Ruochen Ma^{1,2,3,*}, Vedika Khemani^{4,†} and Shengqi Sang^{4,‡}

¹*Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA*

²*Department of Physics, University of California, Santa Barbara, California 93106, USA*

³*Kadanoff Center for Theoretical Physics & Enrico Fermi Institute, The University of Chicago, Chicago, Illinois, USA 60637*

⁴*Department of Physics, Stanford University, Stanford, California 94305, USA*



(Received 19 February 2026; revised 13 May 2026; accepted 6 July 2026; published 10 September 2026)

Quantum phases at zero temperature can be characterized as equivalence classes under local unitary transformations: two ground states within a gapped phase can be transformed into each other via a local unitary circuit. We generalize this circuit-based characterization of phases to systems at finite-temperature thermal equilibrium described by Gibbs states. We construct a channel circuit that approximately transforms one Gibbs state into another provided the two are connected by a path in parameter space along which a certain correlation-decay condition holds. For finite-dimensional systems of linear size L and approximation error ϵ , the locality of the circuit is $\text{Polylog}(\text{Poly}(L)/\epsilon)$. The correlation-decay condition, which we specify, is expected to be satisfied in the interior of many noncritical thermal phases, including those displaying discrete symmetry breaking and topological order. As an application, we show that any system in the same thermal phase as a zero-temperature topological code coherently preserves quantum information for a macroscopically long time, establishing self-correction as a universal property of thermal phases. As part of the proof, we provide explicit encoding and decoding channel circuits to encode information into, and decode it from, a system in thermal equilibrium.

DOI: [10.1103/sh48-wmy4](https://doi.org/10.1103/sh48-wmy4)

I. INTRODUCTION

Over the last several decades, significant progress has been made in understanding zero-temperature quantum phases of gapped local Hamiltonians. This progress covers both the usual symmetry-breaking phases in Landau's picture and many exotic phases with topological or fracton order [1–3]. In this context, a central role is played by a circuit-based definition of phases inspired by quantum information theory: two pure states are in the same phase if there exists a shallow-depth local unitary circuit connecting them [4]. This definition underlies most discussions of the classification and characterization of zero-temperature quantum phases.

Recently, thanks to improved experimental control of devices and other platforms [5], the study of quantum

phases has shifted to open systems described by mixed states. Following the paradigm set by ground-state quantum phase studies, an analogous circuit-based definition has been proposed: two mixed states are in the same phase if they are related by a pair of local quantum channel circuits [6–8], hereafter referred to as two-way connectivity. Based on this, several classification schemes have been proposed [7,9–13], and concrete examples of non-trivial phases have been constructed [14–27].

Gibbs states describing systems in thermal equilibrium are among the most ubiquitous mixed states in nature. It nevertheless remains unclear whether the circuit-based characterization is applicable to Gibbs states and the thermal phases formed by them, where the latter is traditionally defined as a region of parameter space in which the free energy is analytic. To recap, for ground states, the circuit connecting states within a phase is furnished by quasi-adiabatic continuation [28,29], which exists whenever the energy gap does not close. For Gibbs states, an analogous circuit construction that transforms Gibbs states within the same thermal phase is lacking, and it is unclear what the counterpart of an energy gap should be.

In this work, we present a general construction of local channel circuits that transform Gibbs states within the same thermal phase into one another (see Figs. 1 and 2 for

* Contact author: ruochenma@ucsb.edu

† Contact author: vkhemani@stanford.edu

‡ Contact author: sangsq@stanford.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

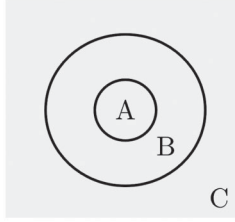


FIG. 1. Annulus tripartition for defining clustering of correlations, illustrated in 2D. A is a small region, B is a region surrounding A , and C denotes the rest of the system.

an illustration). More specifically, for a system of linear size L , we construct a $\text{Polylog}(\text{Poly}L/\epsilon)$ -local channel circuit that approximately transforms one Gibbs state into another with error ϵ , provided that they are connected by a path in the parameter space along which a correlation-decay condition, dubbed stable clustering (briefly described below and formally defined in Definition 2), is satisfied. With an additional standard assumption on the density of states at low energy [30], our construction continues to work in the zero-temperature limit for gapped Hamiltonians.

Colloquially speaking, a Gibbs state $\rho \propto e^{-H}$ satisfies stable clustering if its neighboring Gibbs states of the form $\rho' \propto e^{-H-\delta H}$ all display clustering of correlations when δH is a small perturbation that is a sum of local terms (temperature is absorbed into H and δH). Here, by “clustering of correlations” we mean that the connected correlation function between any local observable O_1 , supported in region A in Fig. 1, and any other, possibly nonlocal observable O_2 , supported in region C , decays

exponentially with the distance between their spatial supports. If the Hamiltonians under consideration satisfy a global (0-form) symmetry, then O_1 and O_2 range only over operators respecting the symmetry. Based on intuition from statistical mechanics, stable clustering is expected to hold in the interiors of a broad class of thermal phases, including those displaying discrete symmetry-breaking order [31] and quantum topological order. Violations of the condition are expected near finite- and zero-temperature phase-transition points due to the presence of long-range correlations.

A notable feature of our construction is that the channel circuit satisfies local reversibility (formally defined in Definition 3), which intuitively means that each channel gate’s action does not disturb long-distance correlations in the state at that time and therefore can be reversed locally. Locally reversible channel circuits were proposed in Ref. [32] as a generalization of local unitary circuits to open quantum systems. Among other consequences, this implies the existence of a reversal circuit that transforms the final state back to the initial state, thereby connecting to the two-way connectivity definition of mixed-state phases. It also guarantees that the circuit preserves sharp features of the state, including symmetries and their anomalies [32].

As an application of the construction, we prove a connection between self-correcting quantum memory and quantum thermal phases. A Hamiltonian system is called a self-correcting quantum memory if it is capable of preserving quantum information for a macroscopically long time when coupled to a low-temperature bath. Prominent examples include toric codes in $D \geq 4$ dimensions [33,34]. At first sight, this property relies on both the system’s Hamiltonian and the specific form of the system–environment interaction. Therefore, it is *a priori* unclear whether (a) self-correction can be inferred from the steady state being a low-temperature Gibbs state alone and (b) if it is a universal property shared by other systems in the same thermal phase. We answer both questions affirmatively. For systems in the same thermal phase as a zero-temperature topological code, we show that any such system acts as a self-correcting quantum memory with an $\exp(L^{\mathcal{O}(1)})$ memory time (Theorem 3). As part of the proof, we construct channels that encode and decode quantum information into and from the system at finite-temperature equilibrium.

For Hamiltonians composed of commuting local terms, we can relax the stable clustering condition to a clustering condition only and obtain an improved construction: if two Gibbs states are connected by a path along which clustering (rather than stable clustering) holds, then there exists an $\mathcal{O}(\text{Polylog}(\text{Poly}L/\epsilon))$ -local Lindbladian evolution that approximately transforms one state into the other (Theorem 4). We expect a similar relaxation to be possible for the noncommuting case as well and leave it for future work.

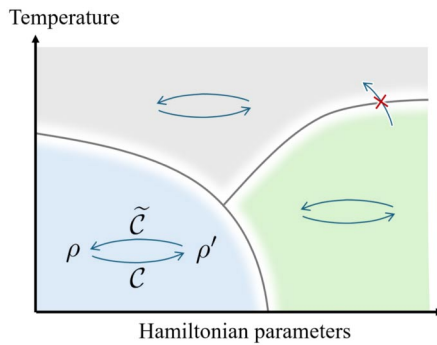


FIG. 2. Sketch of three thermal phases in the temperature-parameter space, separated by phase boundaries (solid lines). In this work, we construct $\text{Polylog}(L)$ -local channel circuits that transform between different Gibbs states within each phase, provided that the states in the phase satisfy a correlation-decay condition defined in Definition 2 (Theorem 1). If the Hamiltonians are commuting, then only a weaker condition, Definition 1, is needed, and the circuit is improved to a continuous-time Lindbladian (Theorem 4). The condition is expected to be satisfied in many ordered phases, including those displaying discrete symmetry breaking and topological order, but violated when approaching phase boundaries.

We note that our results build on several recent advances in quantum information theory. In particular, we use the result, proved in Ref. [35], that the Gibbs state of a local Hamiltonian is approximately Markovian and that, under clustering of correlations, a local perturbation of the Hamiltonian affects only a nearby region of a Gibbs state. Work on Glauber dynamics for the classical Ising model [36], although not phrased explicitly in this manner, also suggests that the symmetry-broken phase of the Ising model possesses some form of circuit connectivity. We further note that a recent work on self-correcting quantum memories studies related questions in the context of stabilizer Hamiltonians, particularly the 4D toric code [37].

Finally, we remark that this work takes a state-centric point of view to study thermal phases. Alternatively, one can take a dynamics-centric point of view by formulating and studying phases of thermalizing dynamics, where Gibbs states emerge as steady states. As exemplified by our result on self-correcting quantum memories (Theorem 2), many properties of the dynamics can already be inferred from its steady state. Nevertheless, there are also properties of the dynamics that are not reflected sharply in the state, with a notable example in finite dimensions being the domain-wall roughening transition of the 3D Ising model [38]. The transition, which occurs at a critical temperature within the Ising model's ordered phase, is a sharp change in the lifetime of a large, traversing domain wall. Interestingly, it is not accompanied by any singular behavior in the Gibbs state, and we expect that Gibbs states below and above the critical temperature can be connected by the channel circuits we construct. More drastic examples, including ones displaying multiple dynamical phase transitions without any thermodynamic transitions, can be found in systems on expander graphs [39–43]. Characterizing these dynamical phases and transitions in a unified framework remains a challenging open problem.

II. SETUP AND DEFINITIONS

We consider a system on a D -dimensional regular lattice, where Λ denotes the set of all sites. For any subset $X \subset \Lambda$, let $|X|$ denote the number of sites in X . For two disjoint subsets X and Y , we define the distance $d(X, Y)$ as the length of the shortest path between them. For an operator O , $\|O\|$ and $\|O\|_1$ denote its operator norm and trace norm, respectively.

We focus on systems with local interactions in thermal equilibrium, described by the Gibbs states:

$$\rho_\beta := \frac{e^{-H_\beta}}{\text{Tr}(e^{-H_\beta})}, \quad H_\beta := \beta \cdot \mathbf{h} = \sum_Z \beta_Z h_Z, \quad (1)$$

where $\mathbf{h} = \{h_Z\}$ is a list of local interaction terms with $\Omega(L^D)$ elements. Each h_Z acts on a region $Z \subset \Lambda$ with $\text{diam}(Z) \leq R$ and $\|h_Z\| \leq 1$. The inverse temperature of ρ_β

is defined as $\|\beta\|_\infty := \sup_Z |\beta_Z|$, in agreement with the literature [35].

We henceforth assume a fixed set of interactions $\mathbf{h}$, chosen to specify the family of Hamiltonians under consideration. Three choices of $\mathbf{h}$ are relevant to this work: (i) all local operators with bounded support, which gives the generic local-Hamiltonian family; (ii) all local operators that are symmetric under a global onsite symmetry, which gives the family of symmetric local Hamiltonians (e.g. Ising-symmetric models); and (iii) a set of mutually commuting local operators, which covers all classical Hamiltonians and stabilizer Hamiltonians.

We use $\mathcal{A}$ to denote the operator algebra generated by $\mathbf{h}$. For a subregion $X \subseteq \Lambda$, we use $\mathcal{A}_X$ to denote the algebra generated by elements of $\text{Tr}_{\bar{X}} \mathcal{A}$. We observe that $\rho_\beta \in \mathcal{A}$ and $\rho_\beta^X \in \mathcal{A}_X$ for any β .

We now formally define the clustering of correlations condition used in this work. For a given operator algebra $\mathcal{A}$ and a state $\rho \in \mathcal{A}$, we define the covariance of two disjoint subsets $X, Y \subset \Lambda$ as

$$\text{Cov}_\rho^{\mathcal{A}}(X, Y) := \sup_{\substack{O_1 \in \mathcal{A}_X: \|O_1\| = 1 \\ O_2 \in \mathcal{A}_Y: \|O_2\| = 1}} |\langle O_1 O_2 \rangle_\rho - \langle O_1 \rangle_\rho \langle O_2 \rangle_\rho|, \quad (2)$$

where $\langle O \rangle_\rho = \text{Tr}(O\rho)$.

Definition 1. (Clustering of correlations). Let $\xi \geq 0$ be a constant and $f(x)$ be a function that grows at most polynomially with x . We say a state $\rho \in \mathcal{A}$ exhibits (ξ, f) clustering if, for every annulus tripartition in Fig. 1, the following holds:

$$\text{Cov}_\rho^{\mathcal{A}}(A, C) \leq f(L) e^{-d(A, C)/\xi}, \quad (3)$$

We remark that this clustering condition is weaker than more common versions in the literature [44] due to the restriction to operators from $\mathcal{A}_A$ and $\mathcal{A}_C$, as well as the particular choice of the A-B-C geometry illustrated in Fig. 1. This makes it applicable to a large class of states with nontrivial orders. For instance, the Ising model below the critical temperature satisfies our definition of clustering, since local operators in $\mathcal{A}_{A/C}$ respect the Ising symmetry and display exponentially decaying correlations, even though nonsymmetric ones have long-range correlations. As another example, topologically ordered ground states on a torus satisfy our notion of clustering, even though they have long-range correlations for non-annulus partitions (see, e.g., [45], which shows that noncontractible loop operators display long-range correlations).

For Gibbs states, we introduce another concept dubbed *stable clustering*:

Definition 2. (Stable clustering). A Gibbs state ρ_β (as defined in Eq. (1)) satisfies (δ, ξ, f) stable clustering if, for

any $\Delta\beta$ with $|\Delta\beta|_\infty \leq \delta$, the Gibbs state $\rho_{\beta+\Delta\beta}$ exhibits (ξ, f) clustering.

Henceforth, when we use the term “stable clustering” without specifying (δ, ξ, f) , it is assumed that $\delta, \xi = \mathcal{O}(1)$ and $f = \text{Poly}L$.

We expect stable clustering to be satisfied in the interior of certain classes of thermal phases. Here, “certain classes” include topologically ordered phases, spontaneous symmetry-breaking (SSB) phases of discrete global symmetries (if $\mathbf{h}$ includes only symmetric terms), and the high-temperature trivial phase. Exceptions should include phases displaying SSB of continuous global symmetries or spatial symmetries; for the former, due to the presence of Goldstone modes leading to power-law-decaying correlations, and for the latter, since the perturbation $\Delta\beta$ can explicitly break spatial symmetry. Stable clustering is also expected to be violated at phase transitions, including both discontinuous (first order) and continuous (higher order) ones.

In open systems, local channel circuits are natural generalizations of unitary circuits. On a lattice Λ , a local channel circuit of depth T is a quantum channel of the form:

$$\mathcal{C} = \mathcal{C}_T \circ \cdots \circ \mathcal{C}_2 \circ \mathcal{C}_1 \quad \text{where } \mathcal{C}_t = \prod_x \mathcal{E}_{t,x}, \quad (4)$$

where each $\mathcal{C}_t$ is called a layer of the circuit, and each $\mathcal{E}_{t,x}$ is called a gate in layer t . Different gates within a layer have non-overlapping spatial support. The range of $\mathcal{C}$, denoted by $\text{rng } \mathcal{C}$, is defined as T times the largest diameter of the supports of the $\mathcal{E}_{t,x}$.

The channel circuits we construct in later sections satisfy a strong property called local reversibility [32], formally defined as

Definition 3. (Local reversibility (LR)). Let $\mathcal{C}$ be a channel circuit as defined above. $\mathcal{C}$ is ϵ_{LR} -locally reversible with respect to an input state ρ if, for each channel gate $\mathcal{E}_{t,x}$, there exists another channel $\tilde{\mathcal{E}}_{t,x}$ with the same spatial support as $\mathcal{E}_{t,x}$ such that

$$\|\tilde{\mathcal{E}}_{t,x} \circ \mathcal{E}_{t,x} \circ \mathcal{C}'[\rho] - \mathcal{C}'[\rho]\|_1 \leq \epsilon_{\text{LR}}, \quad (5)$$

where $\mathcal{C}' = \mathcal{C}_{t-1} \cdots \mathcal{C}_1$ represents the first $(t-1)$ layers of $\mathcal{C}$.

Intuitively, local reversibility means each channel gate’s action does not disturb long-range correlations of the state at that time and can therefore be reversed locally. This gate-by-gate reversibility, among other consequences, implies the existence of a reversal circuit:

$$\tilde{\mathcal{C}} = \tilde{\mathcal{C}}_1 \circ \cdots \circ \tilde{\mathcal{C}}_{T-1} \circ \tilde{\mathcal{C}}_T \quad \text{where } \tilde{\mathcal{C}}_t = \prod_x \tilde{\mathcal{E}}_{t,x}, \quad (6)$$

that reverses $\mathcal{C}$ ’s action on ρ :

$$\|\tilde{\mathcal{C}} \circ \mathcal{C}[\rho] - \rho\|_1 \leq |\mathcal{C}| \cdot \epsilon_{\text{LR}}. \quad (7)$$

Here $|\mathcal{C}|$ denotes the total number of gates within $\mathcal{C}$. A proof of Eq. (7) can be found in Appendix I.

III. CONSTRUCTING CHANNEL CIRCUITS CONNECTING GIBBS STATES

We now state our first result:

Theorem 1. Let ρ_β and $\rho_{\beta'}$ be two Gibbs states connected by a path in β -space along which stable clustering is satisfied. Then there exists a $2\epsilon_{\mathcal{C}}$ -locally reversible channel circuit $\mathcal{C}$ with range $\text{rng } \mathcal{C} = \text{Polylog}(\text{Poly}(L)/\epsilon_{\mathcal{C}})$, such that $\|\rho_{\beta'} - \mathcal{C}[\rho_\beta]\|_1 \leq \epsilon_{\mathcal{C}}$. Here L is the linear system size.

We remark that our construction and the error bound continue to work when one of (or both of) ρ_β and $\rho_{\beta'}$ is (are) zero-temperature gapped ground state(s), if we impose some sparsity conditions on the density of states near ground states, similar to the condition introduced in Ref. [46]. We discuss this point in detail in Sec. III C.

Before proving the theorem in the next two subsections, we review several results on which our proof is based.

The first result relates clustering to the property of *local perturbation that perturbs locally* (LPPL). As its name suggests, LPPL concerns the extent to which a local change in the Hamiltonian modifies the expectation value of distant observables evaluated in the Gibbs state. The theorem is formally stated below. The proof of this result, which closely follows the proof of Theorem 25 in Ref. [44], is included in Appendix A.

Lemma 1. (Clustering implies LPPL. Theorem 25 in Ref. [44], modified). Let $\rho(s)$ be the Gibbs state corresponding to Hamiltonian $H(s) = H(0) + s \cdot h_A$, with h_A supported on a local region A . Then, for any region C non-overlapping with A and any $l < d(A, C)$, one has

$$\begin{aligned} & \|\rho^C(0) - \rho^C(1)\|_1 \\ & \leq \|h_A\| \cdot \left(\sup_{s \in [0,1]} \text{Cov}_{\rho(s)}^A(A_{+l}; C) + 6|A|e^{-\frac{1}{1+v/\pi}l} \right), \end{aligned} \quad (8)$$

where A_{+l} denotes the set of sites whose distance to A is no larger than l , and v is the maximum Lieb-Robinson velocity of $\{H(s)\}_{s \in [0,1]}$.

The second result states that Gibbs states of local Hamiltonians satisfy the so-called approximate Markov property. The property is defined via quantum conditional mutual information (CMI). For any state defined on a tripartite system ABC , CMI is defined as

$$I(A : C|B)_\rho := I(A : BC)_\rho - I(A : B)_\rho, \quad (9)$$

where $I(X : Y)_\rho$ is the quantum mutual information, a measure of correlations between two parties. For Gibbs states on finite-dimensional lattices, we consider the

annulus tripartition (Fig. 1): A is a small simply connected region, B is a buffer region surrounding A , and C is the rest of the system. In this context, CMI measures the locality of correlations: small $I(A : C|B)$ means that A 's correlations with the rest of the system can be well captured by the buffer B —a property called the Markov property. We remark that clustering and the Markov property are not directly related: there are systems displaying clustering but not the Markov property and vice versa. The following theorem states that all Gibbs states satisfy an approximate Markov property locally:

Lemma 2. (Gibbs states are locally Markovian. Ref. [35]). For a Gibbs state ρ_β and any annular-shaped tripartition depicted in Fig. 1, the CMI is bounded as

$$I(A : C|B)_{\rho_\beta} \leq |A| \cdot |C| e^{\kappa + \mu|A| - d(A,C)/\lambda}, \quad (10)$$

where κ , μ , and λ are system-size-independent constants growing at most polynomially with $|\beta|_\infty$.

We note that for any set of β with $\max |\beta|_\infty$ bounded above, one can choose $\kappa, \mu, \lambda = \text{Poly}(\max |\beta|_\infty)$ such that Eq. (10) holds uniformly for all β in the set. For commuting local Hamiltonians a stronger claim holds:

$$I(A : C|B)_{\rho_\beta} = 0 \quad \text{when } \text{dist}(A, C) > R, \quad (11)$$

which is the content of the quantum Hammersley-Clifford theorem [47,48].

The importance of the Markov property to our construction lies in its connection to recoverability. This is stated precisely as follows:

Lemma 3. (Markov property implies recoverability. Ref. [49], modified). For a state ρ defined on a tripartite system $A \cup B \cup C$, there exists a quantum channel $\mathcal{R}_{B \rightarrow AB}^\rho$ acting from B to $A \cup B$ such that

$$\|\rho - \mathcal{R}_{B \rightarrow AB}^\rho \circ \text{Tr}_A(\rho)\|_1 \leq \sqrt{4 \ln 2 \cdot I(A : C|B)_\rho}. \quad (12)$$

An explicit construction of $\mathcal{R}$ is given by the twirled Petz recovery map (see Appendix B for an explicit expression and a proof of the lemma using the main theorem of Ref. [49]).

A. Local interaction variation

We start by constructing a channel that transforms between two Gibbs states whose Hamiltonians differ only in a local region A [see Fig. 3]. Let β satisfy (δ, ξ, f) -stable clustering. We aim to find a channel that changes $\rho = \rho_\beta$ into $\tilde{\rho} = \rho_{\beta + \Delta\beta_A}$ for some simply connected region A and an interaction variation $\Delta\beta_A$ such that $\Delta\beta_A \cdot \mathbf{h}$ is supported strictly within A and $|\Delta\beta_A|_\infty < \delta$.

Let $B = B_1 \cup B_2$ be an annulus-shaped region surrounding A , and C be the rest of the system (Fig. 3).

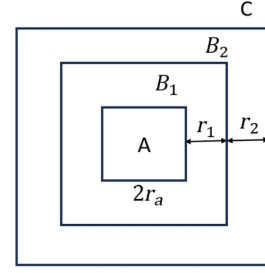


FIG. 3. Geometry for local interaction variation.

We use r_a, r_1, r_2 to denote the radii of A, B_1, B_2 , respectively.

We consider the following channel acting on AB :

$$\mathcal{M} = \mathcal{R}_{B_2 \rightarrow AB}^{\tilde{\rho}} \circ \text{Tr}_{AB_1} \quad (13)$$

where $\mathcal{R}_{B_2 \rightarrow AB}^{\tilde{\rho}}$ is the twirled Petz recovery map using $\tilde{\rho}$ as a reference state (an explicit form of $\mathcal{R}$ can be found in Appendix B). The intuition behind Eq. (13) is to first erase AB_1 from ρ ; by LPPL, the remaining state on B_2C is similar to $\tilde{\rho}^{B_2C}$. One then locally resamples the state within AB_1 according to $\tilde{\rho}$ using the recovery map in Eq. (12).

As we show, $\mathcal{M}$ approximately transforms ρ into $\tilde{\rho}$, with an error bounded by

$$\begin{aligned} \|\tilde{\rho} - \mathcal{M}(\rho)\|_1 &\leq \epsilon_{\mathcal{M}} \\ \epsilon_{\mathcal{M}} &= \text{Poly}(L)(e^{\kappa + \mu|AB_1| - r_2/\lambda} + e^{-r_1/\xi'}), \end{aligned} \quad (14)$$

where $\xi' = \xi + 1 + v/\pi$, with v being the maximum Lieb-Robinson velocity [44] of the Hamiltonians in $\{H_{\beta+s\Delta\beta_A}\}_{s \in [0,1]}$. The other constants κ, μ , and λ are as in Eq. (10).

The error bound Eq. (14) is derived by invoking Lemmas 1–3. Using the triangle inequality (sub- and superscripts of the recovery map $\mathcal{R}$ are omitted):

$$\|\tilde{\rho} - \mathcal{M}[\rho]\|_1 \leq \|\tilde{\rho} - \mathcal{R}[\tilde{\rho}^{B_2C}]\|_1 + \|\mathcal{R}[\tilde{\rho}^{B_2C}] - \mathcal{R}[\rho^{B_2C}]\|_1. \quad (15)$$

The approximate Markov property of Gibbs states (Lemmas 2 and 3) guarantees that the first term is bounded by the $e^{\kappa + \mu|AB_1| - r_2/\lambda}$ term. Further, since ρ satisfies stable clustering, Lemma 1 implies ρ and $\tilde{\rho}$ are approximately indistinguishable on B_2C :

$$\begin{aligned} \|\tilde{\rho}^{B_2C} - \rho^{B_2C}\|_1 &\leq \delta \cdot \left(f(L) e^{-\frac{1}{\xi}(r_1-l)} + 6|A| e^{-\frac{1}{1+v/\pi}l} \right) \quad \forall l < r_1 \\ &= \text{Poly}(L) e^{-r_1/\xi'}, \end{aligned} \quad (16)$$

where the last equality is obtained by letting $l = (1 + \frac{\xi}{1+v/\pi})^{-1} r_1$, which makes the two exponential

factors equal. The corresponding ξ' is $\xi' = \xi + 1 + v/\pi$. This finishes the proof of Eq. (14).

We now consider the trade-off between $\mathcal{M}$'s locality and the approximation error. To ensure $\|\mathcal{M}[\rho] - \tilde{\rho}\|_1 \leq \epsilon_{\mathcal{M}}$, it suffices to choose

$$\begin{aligned} r_1 &= r_a \geq \xi' \ln(\text{Poly}(L)/\epsilon_{\mathcal{M}}) \\ r_2 &\geq \kappa\lambda + \mu\lambda(4r_1)^D + \lambda \ln(\text{Poly}(L)/\epsilon_{\mathcal{M}}). \end{aligned} \quad (17)$$

If we further assume the inverse temperature $|\beta|_{\infty}$ is at most $\log(\text{Poly}(L)/\epsilon_{\mathcal{M}})$, we conclude that

$$r_a, r_1, r_2 = \text{Polylog}(\text{Poly}L/\epsilon_{\mathcal{M}}). \quad (18)$$

Before proceeding, we note that the channel $\mathcal{M}$'s action on ρ can be approximately reversed by

$$\tilde{\mathcal{M}} := \mathcal{R}_{B_2 \rightarrow AB}^{\rho} \circ \text{Tr}_{AB_1}, \quad (19)$$

which is of the same form as $\mathcal{M}$ but replaces $\tilde{\rho}$ by ρ in the recovery map $\mathcal{R}$. Following the same proof, we have $\|\tilde{\mathcal{M}}[\tilde{\rho}] - \rho\|_1 \leq \epsilon_{\mathcal{M}}$, and therefore

$$\|\tilde{\mathcal{M}} \circ \mathcal{M}[\rho] - \rho\|_1 \leq 2\epsilon_{\mathcal{M}}. \quad (20)$$

B. Global interaction variation

Building on local interaction variation, we now describe a layered channel circuit that maps between two Gibbs states ρ_{β} and $\rho_{\beta'}$ connected by a path satisfying stable clustering. The strategy is to make small, local changes to the Hamiltonian each time, in a carefully chosen order so that the overall circuit implementing the global interaction variation has bounded range.

We first observe that it suffices to construct a circuit $\mathcal{G}$ that maps ρ_{β} to $\rho_{\beta+\Delta\beta}$ with $|\Delta\beta|_{\infty} \leq \delta$ (δ as in the stable clustering definition, Definition 2) and then repeat the procedure n times. This is because we can discretize the path connecting β to β' into a sequence of length $n \sim |\beta - \beta'|_{\infty}/\delta$ (δ as in Definition 2) $\beta_0 - \beta_1 - \dots - \beta_n$, such that $\beta_0 = \beta, \beta_n = \beta'$ and $|\beta_m - \beta_{m-1}|_{\infty} \leq \delta$ ($0 < m \leq n$).

To this end, we begin by partitioning the system into hypercubic blocks in D spatial dimensions, with each block having diameter $2r_a$. Two adjacent blocks overlap along their boundaries in a region of width R (recall that R is the range of interaction terms in $\mathbf{h}$). We denote each block region by A_x , with x being the center of the block. There are $N = L^D/(2r_a - R)^D$ blocks in total.

This accordingly leads to a partition of the Hamiltonian terms: $\beta \cdot \mathbf{h} = \sum_x \beta_x \cdot \mathbf{h}_x$, where $\mathbf{h}_x$ contains only h_Z 's supported entirely within A_x and β_x are their coefficients. Due to the width- R overlap, each h_Z is contained in at least one A_x . If an h_Z is in more than one block, it is assigned to only one of them.

We observe that for any ordering of blocks $\{A_{x_1}, A_{x_2}, \dots, A_{x_N}\}$ we can turn ρ_{β} into $\rho_{\beta+\Delta\beta}$ by varying interactions on different blocks sequentially, using the scheme described in the previous section. More specifically, for $i = 0, 1, \dots, N$ let

$$\rho^{(i)} := \rho_{\beta+\Delta\beta_i} \quad \text{where } \Delta\beta_i \cdot \mathbf{h} := \sum_{j \leq i} \Delta\beta_{x_j} \cdot \mathbf{h}_{x_j}, \quad (21)$$

and let $\mathcal{M}^{(i)}$ be the channel defined in Eq. (13) with $\tilde{\rho}$ set to $\rho^{(i+1)}$. Then one can show that the error of the global interaction variation is bounded by the sum of the errors in the local steps (see Appendix D for a detailed derivation):

$$\begin{aligned} \epsilon_{\mathcal{G}} &:= \|\mathcal{G}[\rho_{\beta}] - \rho_{\beta+\Delta\beta}\|_1 \\ \epsilon_{\mathcal{G}} &\leq \sum_{i=0}^{N-1} \|\mathcal{M}^{(i)}[\rho^{(i)}] - \rho^{(i+1)}\|_1 \leq N \cdot \epsilon_{\mathcal{M}}, \end{aligned} \quad (22)$$

where $\mathcal{G} := \mathcal{M}^{(N)} \circ \dots \circ \mathcal{M}^{(1)}$.

We devise an optimal ordering of blocks so that the resulting $\mathcal{G}$ has minimal depth and range. We note that $\mathcal{M}^{(i)}$ and $\mathcal{M}^{(j)}$ are non-overlapping and can be implemented in parallel if x_i and x_j are at least $2r_b := 2(r_a + r_1 + r_2)$ apart. Therefore, a single layer can accommodate $(L/r_b)^D$ non-overlapping gates by placing gates on an r_b -spaced grid (see Fig. 4). In this way, $\mathcal{G}$ can be organized into a circuit with at most $N \cdot (L/r_b)^{-D} \approx (r_b/r_a)^D$ layers.

The total circuit $\mathcal{C} = \circ_{m=1}^n \mathcal{G}_m$, which transforms ρ_{β} to $\rho_{\beta'}$ along the path $\rho_{\beta_1}, \dots, \rho_{\beta_n}$, has n times the depth and approximation error of each $\mathcal{G}_m$. Each $\mathcal{G}_m$ transforms $\rho_{\beta_{m-1}}$ to ρ_{β_m} following the construction above. Combining Eqs. (22) and (18), we conclude that having

$$\text{rng } \mathcal{C} = \text{Polylog}(\text{Poly}L/\epsilon_{\mathcal{C}}) \quad (23)$$

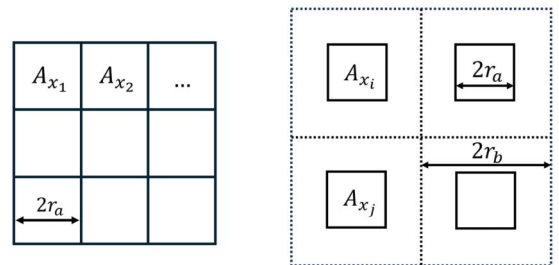


FIG. 4. Illustration of global interaction variation. Left: the system is partitioned into a set of diameter $2r_a$ blocks $\{A_{x_i}\}_i$. Adjacent blocks have a width R overlap on their boundary, which is omitted in the figure. Right: the gate $\mathcal{M}^{(i)}$ used for varying interaction on A_{x_i} supports within the dashed box containing A_{x_i} . Two gates $\mathcal{M}^{(i)}$ and $\mathcal{M}^{(j)}$ can be placed in a single layer and implemented in parallel if x_i and x_j are at least $2r_b$ apart.

is sufficient to ensure $\|\mathcal{C}[\rho_\beta] - \rho_{\beta'}\|_1 \leq \epsilon_C$, provided the inverse temperatures are bounded above by $|\beta|_\infty, |\beta'|_\infty \leq \log(\text{Poly}L/\epsilon_C)$. In addition, as we prove in Appendix I, the constructed $\mathcal{C}$ is $2\epsilon_C$ -LR with respect to ρ_β . This completes the proof of Theorem 1.

C. Zero-temperature limit

In this section we discuss how our construction can be applied when the initial (or equivalently the target) Gibbs state is a maximally mixed state in the ground-state subspace of some H_β [50], denoted by $\rho_{\text{g.s.}}$. We assume $|\beta|_\infty = 1$ in this section without loss of generality.

When there is an energy gap above the ground-state subspace, $\rho_{\text{g.s.}}$ can be well approximated by $\rho_{s\beta}$ for large enough s (low enough temperature) in many physical Hamiltonians. Following the argument in Ref. [30], if we assume that low-lying excitations of H_β are all closed p -dimensional manifolds (e.g., particles, loops, membranes, etc.), then the number of low-lying states grows at most polynomially with the system size, leading to the approximation error (see Appendix F for a detailed argument and derivations):

$$\|\rho_{\text{g.s.}} - \rho_{s\beta}\|_1 = \text{Poly}(L)e^{-s\Delta}, \quad (24)$$

where $\Delta = \mathcal{O}(1)$ is H_β 's energy gap above the ground-state subspace.

Hence, to guarantee an ϵ approximation error, we need s to be as large as $\log(\text{Poly}L/\epsilon)$. Now we can follow the methods in previous sections and construct a channel circuit $\mathcal{C}$ of range $\text{Polylog}(\text{Poly}L/\epsilon)$ that transforms $\rho_{s\beta}$ to the final state $\rho_{\beta'}$, with error also bounded by ϵ . This leads to an error bound:

$$\|\mathcal{C}[\rho_{\text{g.s.}}] - \rho_{\beta'}\|_1 \leq \|\mathcal{C}[\rho_{s\beta}] - \rho_{\beta'}\|_1 + \|\rho_{s\beta} - \rho_{\text{g.s.}}\|_1 \leq 2\epsilon. \quad (25)$$

Following an analogous procedure, we can handle the case in which $\rho_{\beta'}$ approaches the zero-temperature limit.

IV. THERMAL PHASES OF SELF-CORRECTING QUANTUM MEMORIES

As an application of Theorem 1, in this section we show that quantum self-correction is a universal property shared by all states within the same thermal phase as a zero-temperature topological quantum code. More specifically, if $\mathcal{C}$ is the constructed circuit (following the recipe in the previous section) connecting a quantum code to another Gibbs state ρ_β , we prove that $\mathcal{C}$ encodes a pure codeword state into a mixed state that is locally indistinguishable (in the sense of Definition 4) but globally distinguishable from ρ_β . Due to the local indistinguishability from ρ_β , the encoded state is long-lived under any quasi-local dynamics with ρ_β as a steady state.

We start by formulating the self-correcting property formally. We model the environment's influence on the system, as well as the system's internal evolution, jointly by a quasi-local Lindbladian super-operator of the form:

$$\mathcal{L} = \sum_X \mathcal{L}_X, \quad (26)$$

such that each $\mathcal{L}_X$ satisfies $\|\mathcal{L}_X\|_{1 \rightarrow 1} \leq 1$. The quasi-locality of each $\mathcal{L}_X$ means there exists an approximating family of diameter- l -support super-operators $\{\mathcal{L}_X^l\}$, such that $\|\mathcal{L}_X - \mathcal{L}_X^l\|_{1 \rightarrow 1} \leq e^{-\alpha l}$ with a constant $\alpha > 0$ [51]. Since the system's Gibbs state ρ_β emerges as the steady state of system-environment interaction, it follows that $\mathcal{L}[\rho_\beta] = 0$. Here we assume the stronger property $\mathcal{L}_X[\rho_\beta] = 0$, which means ρ_β is a frustration-free steady state of $\mathcal{L}$. Concrete instances of $\mathcal{L}$ satisfying all properties above have recently been constructed in Ref. [52].

We say a Lindbladian $\mathcal{L}$ preserves $\log_2 K$ qubits of quantum information for a memory time t with error ϵ if there exists an encoding channel $\text{En}: \mathcal{S}(\mathbb{C}^K) \mapsto \mathcal{S}(\mathcal{H})$ (here $\mathcal{H}$ denotes the system's Hilbert space, and $\mathcal{S}(x)$ denotes the set of density operators defined on a Hilbert space x) and a decoding channel $\text{De}: \mathcal{S}(\mathcal{H}) \mapsto \mathcal{S}(\mathbb{C}^K)$ such that:

$$\|\text{De} \circ e^{\tau \mathcal{L}} \circ \text{En}(\omega) - \omega\|_1 \leq \epsilon \quad \forall \omega \in \mathcal{S}(\mathbb{C}^K), \tau \leq t. \quad (27)$$

We call a local Hamiltonian's zero-temperature Gibbs state Π (which, as we recall, is the maximally mixed state in its ground-state subspace V) a diameter- ℓ quantum code if $\Pi O \Pi \propto \Pi$ for any operator O supported on a region with a diameter no larger than ℓ [53]. Topological codes, i.e., quantum codes whose code states display quantum topological order, have a diameter of at least $L/2$.

As a crucial ingredient of the proof that self-correction is a universal property, the following theorem shows that whether a local dynamics can preserve quantum information may be inferred from the properties of its steady state.

Theorem 2. Suppose a state ρ satisfies $\|\mathcal{C}(\Pi) - \rho\|_1 \leq \epsilon_C$ for a diameter- ℓ code Π and a channel circuit $\mathcal{C}$. Further, suppose $\mathcal{C}$ is ϵ_{LR} -locally reversible with respect to Π and satisfies $\text{rng } \mathcal{C} \leq \ell/4$. Then any quasi-local Lindbladian $\mathcal{L}$ having ρ as a frustration-free steady state preserves $\log_2(\text{rank } \Pi)$ qubits of information for time t with error

$$\epsilon_t = \text{Poly}(L)(\epsilon_{LR} + \epsilon_C + e^{-\alpha \ell/2}) \cdot (t + 1). \quad (28)$$

We remark that we do not assume $\mathcal{L}$ to satisfy quantum detailed balance, and therefore the theorem is applicable to Lindbladian dynamics beyond thermalization.

The proof of Theorem 2 utilizes *local indistinguishability* (LI) of states within a topological code space, formally defined below.

Definition 4. (Local indistinguishability (LI) for mixed states [32]). Two states ρ and σ are (r, ϵ_{LI}) -locally indistinguishable if, for any simply connected A with $\text{diam}(A) \leq r$, there exist channels $\mathcal{D}, \mathcal{D}'$ acting outside A such that $\|\mathcal{D}[\rho] - \sigma\|_1 \leq \epsilon_{LI}$ and $\|\mathcal{D}'[\sigma] - \rho\|_1 \leq \epsilon_{LI}$.

It is straightforward to see that if ρ and σ are LI, then $\rho^A \approx \sigma^A$ for any A with $\text{diam}(A) \leq r$ since the channel $\mathcal{D}$ (or $\mathcal{D}'$) does not act on A , hence the name “local indistinguishability” [54]. If V is a code space for a diameter- ℓ quantum code, then any two codeword states are $(\ell, 0)$ -LI (see Appendix H for a proof), a conclusion known as the cleaning lemma [55] in quantum coding theory.

The following lemma (whose proof can be found in Appendix I) shows that local indistinguishability is preserved under a locally reversible circuit.

Lemma 4. Suppose ρ and σ are $(\ell, 0)$ -LI and the channel circuit $\mathcal{C}$ is ϵ_{LR} -LR with respect to ρ . Then:

1. $\mathcal{C}$ is $(2|\mathcal{C}| \epsilon_{LR} + \epsilon_{LR})$ -LR with respect to σ with the same set of reversal gates, and
2. $\mathcal{C}[\rho]$ and $\mathcal{C}[\sigma]$ are $(\ell - 2\text{rng } \mathcal{C}, (2|\mathcal{C}|^2 + |\mathcal{C}|)\epsilon_{LR})$ -LI, where $|\mathcal{C}|$ denotes the number of gates in $\mathcal{C}$.

We are now ready to prove Theorem 2.

Following the definition in Eq. (27), we identify the subspace spanned by Π as $\mathbb{C}^K$ and take the encoder and decoder to be $\text{En} = \mathcal{C}$ and $\text{De} = \tilde{\mathcal{C}}$ [recall that $\tilde{\mathcal{C}}$ is the reversal circuit of $\mathcal{C}$, defined in Eq. (6)]. The amount of logical error at time t is upper-bounded by

$$\begin{aligned} \epsilon_t &:= \|\text{De} \circ e^{t\mathcal{L}} \circ \text{En}[\omega] - \omega\|_1 \\ &= \|\tilde{\mathcal{C}} e^{t\mathcal{L}} \mathcal{C}[\omega] - \omega\|_1 \\ &\leq \|\tilde{\mathcal{C}} \mathcal{C}[\omega] - \omega\|_1 + \|\tilde{\mathcal{C}} e^{t\mathcal{L}} \mathcal{C}[\omega] - \tilde{\mathcal{C}} \mathcal{C}[\omega]\|_1 \\ &\leq \|\tilde{\mathcal{C}} \mathcal{C}[\omega] - \omega\|_1 + \|e^{t\mathcal{L}} \mathcal{C}[\omega] - \mathcal{C}[\omega]\|_1 \end{aligned} \quad (29)$$

It suffices to bound the two terms on the last line separately.

The first term is bounded since $\mathcal{C}$ is locally reversible with respect to ω , as shown below. Since ω and Π are states within the quantum code space, they are $(\ell, 0)$ -LI. According to Lemma 4, $\mathcal{C}$ is $(2|\mathcal{C}| + 1)\epsilon_{LR}$ -LR with respect to ω , which implies that the first term is bounded by $(2|\mathcal{C}|^2 + |\mathcal{C}|)\epsilon_{LR} = \text{Poly}(L) \cdot \epsilon_{LR}$.

Now we show that $\mathcal{C}[\omega]$ is an approximate frustration-free steady state of the Lindbladian $\mathcal{L}$, leading to a bound on the second term. The underlying intuition is simple: since $\mathcal{C}$ preserves local indistinguishability, $\mathcal{C}[\omega]$ and $\mathcal{C}[\Pi]$ remain LI. Therefore, any Lindbladian term $\mathcal{L}_X$ having the latter as a steady state must also leave the former almost unchanged.

We now turn this intuition into a proof with an explicit error bound. By Lemma 4, $\mathcal{C}[\omega]$ and $\mathcal{C}[\Pi]$ are $(\ell/2, (2|\mathcal{C}|^2 + |\mathcal{C}|)\epsilon_{LR})$ -LI, which by definition implies there exists a channel $\mathcal{D}$, which can be chosen to be supported outside any diameter- $\ell/2$ region, such that

$\|\mathcal{D} \circ \mathcal{C}[\Pi] - \mathcal{C}[\omega]\|_1 \leq (2|\mathcal{C}|^2 + |\mathcal{C}|)\epsilon_{LR}$. We therefore obtain the following series of approximations (the notation $A \stackrel{\epsilon}{\sim} B$ means $\|A - B\|_1 \leq \epsilon \cdot \text{Poly}(L)$):

$$\begin{aligned} \mathcal{L}_X \circ \mathcal{C}[\omega] &\stackrel{\epsilon_{LR}}{\sim} \mathcal{L}_X \circ \mathcal{D} \circ \mathcal{C}[\Pi] \stackrel{\epsilon_C}{\sim} \mathcal{L}_X \circ \mathcal{D}[\rho] \stackrel{e^{-al}}{\sim} \\ \mathcal{L}_X^l \circ \mathcal{D}[\rho] &= \mathcal{D} \circ \mathcal{L}_X^l[\rho] \stackrel{e^{-al}}{\sim} \mathcal{D} \circ \mathcal{L}_X[\rho] = 0 \end{aligned} \quad (30)$$

In the equality step, we use the fact that we can choose $\mathcal{D}$ so that it is non-overlapping with $\mathcal{L}_X^l$, as long as $l \leq \ell/2$.

Therefore $\|\mathcal{L} \circ \mathcal{C}[\omega]\|_1 \leq \sum_X \|\mathcal{L}_X \circ \mathcal{C}[\omega]\|_1 \leq \text{Poly}(L)(\epsilon_{LR} + \epsilon_C + e^{-al})$, which leads to

$$\begin{aligned} \|e^{t\mathcal{L}} \mathcal{C}[\omega] - \mathcal{C}[\omega]\|_1 &\leq t \cdot \|\mathcal{L} \circ \mathcal{C}[\omega]\|_1 \\ &\leq t \cdot \text{Poly}(L)(\epsilon_{LR} + \epsilon_C + e^{-al}). \end{aligned} \quad (31)$$

Here the first inequality uses the fact that $\|e^{t\mathcal{L}}[\rho] - \rho\|_1 \leq t\|\mathcal{L}[\rho]\|_1$, which we derive in Appendix J. Setting $l = \ell/2$ completes the proof of Theorem 2.

Theorem 3. (Self-correction is a property of thermal phases). Suppose ρ_β is connected to a zero-temperature diameter- $L/2$ quantum code Π by a path along which stable clustering is satisfied. Then any quasi-local Lindbladian $\mathcal{L}$ having ρ_β as a frustration-free steady state preserves $\log_2(\text{rank } \Pi)$ qubits of quantum information for time t with error:

$$\epsilon_t \leq (t+1) \cdot \exp(-L^{\mathcal{O}(1)}) \quad (32)$$

for sufficiently large L .

This theorem is obtained by combining Theorems 1 and 2, as detailed below.

We take $\mathcal{C}$ as the circuit in Theorem 1 connecting Π to ρ_β . Recall that the construction is such that $\mathcal{C}$ is $2\epsilon_C$ -locally reversible with respect to Π when the approximation error satisfies $\|\mathcal{C}[\Pi] - \rho_\beta\| \leq \epsilon_C$.

Since $\text{rng } \mathcal{C} = \text{Polylog}(\text{Poly}(L)/\epsilon_C)$, we can always find positive constants c_1, c_2 such that $\text{rng } \mathcal{C} \leq \log^{c_2}(L^{c_1}/\epsilon_C)$ holds for sufficiently large L . Letting $\text{rng } \mathcal{C} = L/8$ leads to an error bound

$$\epsilon_C \leq \exp(-(L/8)^{c_2^{-1}}) \cdot L^{c_1} \leq \exp(-L^{c_2^{-1}+\epsilon}), \quad (33)$$

where, for any positive ϵ , the second inequality holds when L is large enough. Now Theorem 3 straightforwardly follows from Theorem 2. We remark that in our construction c_2 is strictly larger than 1; therefore, the error term in Theorem 3 follows a stretched-exponential decay in L .

V. CONTINUOUS-TIME CONSTRUCTION FOR COMMUTING HAMILTONIANS

In this section, we show that for the special case of commuting Hamiltonians (i.e., terms in $\mathbf{h}$ commute with each other), we can relax the stable clustering condition to the clustering condition only (i.e., Definition 2 with $\delta = 0$) and obtain a continuous-time Lindbladian evolution connecting two Gibbs states within the same phase.

Theorem 4. Suppose terms in $\mathbf{h} = \{h_Z\}_Z$ are commuting. Let $\{\beta(s), s \in [0, 1]\}$ be a single-parameter family of couplings satisfying $|\partial_s \beta_{s,Z}| \leq 1$ for all s and Z , and $\{\rho_s \propto \exp(-\beta_s \cdot \mathbf{h})\}_s$ be the corresponding family of Gibbs states. Then, for any $r \geq 0$, there exists an s -dependent Lindbladian $\mathcal{L}_s = \sum_Z \mathcal{L}_{s,Z}$ such that:

$$\|\partial_s \rho_s - \mathcal{L}_s[\rho_s]\|_1 \leq \sum_Z \text{Cov}_{\rho_s}^A(Z, \overline{Z_{+r}}), \quad (34)$$

where Z_{+r} denotes lattice sites whose distance to Z is no more than r . Each $\mathcal{L}_{s,Z}$ acts on $Z_{+(r+R)}$ and satisfies $\|\mathcal{L}_{s,Z}\|_{1 \rightarrow 1} \leq 4$.

According to the theorem, we notice that if (ξ, f) -clustering as defined in Definition 1 is satisfied by each of the ρ_s along the path $s \in [0, 1]$, then the Lindbladian evolution approximately transforms ρ_0 into ρ_1 , with an error bounded as

$$\epsilon_{\mathcal{L}} := \|\rho_1 - \mathcal{T} e^{\int \mathcal{L}_{ds}}[\rho_0]\|_1 \leq \text{Poly}(L) e^{-r/\xi}, \quad (35)$$

which can be rewritten as a bound on the range r of the Lindbladian:

$$r \leq \xi \cdot \log\left(\frac{\text{Poly}(L)}{\epsilon_{\mathcal{L}}}\right). \quad (36)$$

We now prove Theorem 4. Expanding the derivative of ρ_s , we get:

$$\begin{aligned} \partial_s \rho_s &= \sum_Z (\partial_s \beta_{s,Z}) \cdot \partial_{\beta_Z} \rho_s \\ &= \sum_Z (\partial_s \beta_{s,Z}) \cdot \rho_s \cdot (\langle h_Z \rangle_s - h_Z), \end{aligned} \quad (37)$$

where $\langle h_Z \rangle_s = \text{tr}(h_Z \rho_s)$. We assume $\partial_s \beta_{s,Z} \geq 0$ for each Z without loss of generality (otherwise we can redefine h_Z as $-h_Z$).

We make the following key observation: each term of the derivative—e.g., the one corresponding to h_Z —is proportional to the difference between two Gibbs states whose Hamiltonians differ only on Z :

$$\begin{aligned} \rho_s \cdot (\langle h_Z \rangle_s - h_Z) &= (\lambda_Z - \langle h_Z \rangle_s) \cdot (\tilde{\rho}_{s,Z} - \rho_s) \\ \text{where } \tilde{\rho}_{s,Z} &:= \frac{\lambda_Z - h_Z}{\lambda_Z - \langle h_Z \rangle_s} \rho_s \end{aligned} \quad (38)$$

Here λ_Z denotes the largest eigenvalue of h_Z . It is straightforward to check that $\tilde{\rho}_{s,Z}$ is a valid density matrix and can be viewed as a Gibbs state corresponding to the Hamiltonian:

$$\tilde{H}_{s,Z} = \beta_s \cdot \mathbf{h} - \log(\lambda_Z - h_Z + 0_+), \quad (39)$$

which is also commuting.

This observation motivates defining the Lindbladian as

$$\begin{aligned} \mathcal{L}_s &= \sum_Z \mathcal{L}_{s,Z} \quad \text{where} \\ \mathcal{L}_{s,Z} &:= (\partial_s \beta_{s,Z}) \cdot (\lambda_Z - \langle h_Z \rangle_s) \cdot (\mathcal{M}_{s,Z} - \mathcal{I}) \end{aligned} \quad (40)$$

with $\mathcal{I}$ being the identity channel and $\mathcal{M}_{s,Z}$ a channel to be specified soon. It is straightforward to check that $\mathcal{L}_{s,Z}$ is a valid Lindbladian, i.e., $e^{t\mathcal{L}_{s,Z}}$ is a quantum channel for any t . Further, since $|\partial_s \beta_{s,Z}| \leq 1$ and $|\lambda_Z - \langle h_Z \rangle_s| \leq 2\|h_Z\| \leq 2$, we have

$$\|\mathcal{L}_{s,Z}\|_{1 \rightarrow 1} \leq 2\|\mathcal{M}_{s,Z} - \mathcal{I}\|_{1 \rightarrow 1} \leq 4. \quad (41)$$

The Lindbladian above leads to an approximation error:

$$\begin{aligned} \|\partial_s \rho_s - \mathcal{L}_s[\rho_s]\|_1 &\leq \sum_Z \|\partial_s \beta_{s,Z} \partial_{\beta_Z} \rho_s - \mathcal{L}_{s,Z}[\rho_s]\|_1 \\ &\leq \sum_Z |\lambda_Z - \langle h_Z \rangle_s| \cdot \|\tilde{\rho}_{s,Z} - \mathcal{M}_{s,Z}[\rho_s]\|_1. \end{aligned} \quad (42)$$

Thus it remains to choose $\mathcal{M}_{s,Z}$ that approximately transforms ρ_s into $\tilde{\rho}_{s,Z}$. Since the Hamiltonians of ρ_s and $\tilde{\rho}_{s,Z}$ differ only on Z , we can use the same strategy as in Sec. III A. Letting $B_1 = Z_{+r} \setminus Z$, $B_2 = Z_{+(r+R)} \setminus Z_{+r}$, and $C = \overline{Z_{+(r+R)}}$, we define:

$$\mathcal{M}_{s,Z} = \mathcal{R}_{B_2 \rightarrow ZB}^{\tilde{\rho}_{s,Z}} \circ \text{Tr}_{ZB_1}, \quad (43)$$

where $B = B_1 B_2$ and $\mathcal{R}_{B_2 \rightarrow ZB}^{\tilde{\rho}_{s,Z}}$ is the Petz recovery map. Since $\tilde{\rho}$ is a commuting Hamiltonian's Gibbs state and $\text{dist}(Z, B_2) \geq R$, we know $I_{\tilde{\rho}}(Z: B_2 | B_1) = 0$, which leads to $\tilde{\rho}_{s,Z} = \mathcal{R}_{B_2 \rightarrow ZB}^{\tilde{\rho}_{s,Z}}[\tilde{\rho}_{s,Z}^{B_2 C}]$. Therefore:

$$\begin{aligned}
 \|\tilde{\rho}_{s,Z} - \mathcal{M}_{s,Z}[\rho_s]\|_1 &\leq \|\rho_s^{B_2C} - \tilde{\rho}_{s,Z}^{B_2C}\|_1 = \sup_{O \in \mathcal{A}_{B_2C}, \|O\|=1} |\text{Tr}(O\rho_s - O\tilde{\rho}_{s,Z})| \\
 &= \frac{1}{|\lambda_Z - \langle h_Z \rangle_s|} \sup_{O \in \mathcal{A}_{B_2C}, \|O\|=1} |\langle Oh_Z \rangle_s - \langle O \rangle_s \langle h_Z \rangle_s| \leq \frac{1}{|\lambda_Z - \langle h_Z \rangle_s|} \text{Cov}_{\rho_s}^A(Z, \overline{Z_{+r}}), \quad (44)
 \end{aligned}$$

which completes the proof.

VI. SUMMARY AND OUTLOOK

To summarize, in this work we provide a circuit-based characterization of Gibbs states within a thermal phase. We formulate a correlation-decay condition dubbed stable clustering (Definition 2), which is expected to be satisfied for Gibbs states within a vast class of thermal phases. We prove that there exists a local quantum channel circuit that transforms one Gibbs state into another, provided they are connected by a path that satisfies stable clustering. For the case of commuting local Hamiltonians, we are able to relax stable clustering to a weaker condition (Definition 1) and improve the circuit to a continuous-time evolution. As an application of the construction, we establish that if a system is in the same thermal phase as a zero-temperature topological code, then it can coherently store quantum information for a macroscopically long time while in contact with the thermal bath.

We end this work with some discussion and open questions:

- (i) Our constructions have room for potential improvements. In light of Theorem 4, it is natural to ask whether one can relax the stability part of the clustering condition for noncommuting Hamiltonians as well and obtain a continuous-time Lindbladian. In addition, we ask whether the transforming circuit/Lindbladian can be made exact ($\epsilon_C = 0$) while still preserving locality. The latter question is motivated by the exact quasi-adiabatic continuation for gapped ground states (see e.g., Ref. [29]), which exactly transforms one ground state into another while having a transient Hamiltonian with superpolynomially decaying tails.
- (ii) Rigorously establishing the relationship between (stable) clustering and other characterizations of thermal phases, e.g., analyticity of the free energy, remains an open problem. Reference [56] shows that analyticity of the free energy at high temperature implies decay of correlations for local observables in quantum Gibbs states. Extending this connection to low temperatures and to correlations of nonlocal observables (as in the annulus tripartition) remains open. It is also desirable to obtain (stable) clustering for important models, especially at low temperatures. In Appendix G, we provide a physical argument suggesting that stable clustering

fails for the 2D toric code but holds for the 4D toric code near the zero-temperature limit. Turning this into a rigorous proof is desirable.

- (iii) Our current construction of local interaction variation [Eq. (13)], which is an erase-and-resample procedure, is a choice made for analytical tractability rather than a unique requirement. A physically more natural alternative is a short-time evolution under a detailed-balanced quasi-local Lindbladian of the type constructed in Ref. [35], which has the target state $\tilde{\rho}$ as a steady state. One expects such a dissipative evolution to smoothly deform ρ into $\tilde{\rho}$ within AB_1 over a short time. Obtaining a rigorous bound on the time required for this smooth deformation to reach a given accuracy, however, remains beyond the reach of currently available techniques, and we leave it as an interesting open question.
- (iv) The circuit construction under additional symmetry constraints deserves further study. When the Hamiltonian has a global (0-form) symmetry, a natural question is whether the circuit can be made strongly symmetric—our current construction is not, because a partial trace appears in the local interaction variation [Eq. (13)]. In Appendix C we provide evidence that, for onsite $\mathbb{Z}_2$ symmetry, local interaction variation can be made strongly symmetric. Making this statement rigorous and extending it to continuous symmetries is desirable. Furthermore, when a continuous symmetry is present and spontaneously broken at low temperatures, Goldstone modes produce algebraic spatial correlations, so stable (exponential) clustering fails. Whether two Gibbs states in the ordered regime can still be connected by a circuit—even without requiring strong symmetry—and how its range scales with the system size remain open questions.

ACKNOWLEDGMENTS

We thank Thiago Bergamaschi, Matthew P. A. Fisher, Tarun Grover, Tim Hsieh, Michael Levin, Yaodong Li, David Long, Benedikt Placke, Tibor Rakovszky, Mehdi Soleimanifar, Charles Stahl, Chong Wang and Cenke Xu for helpful discussions. S. S. was supported by the SITP postdoctoral fellowship at Stanford University. V. K. acknowledges support from the Packard Foundation through a Packard Fellowship in Science and Engineering and the Office of Naval Research Young

Investigator Program (ONR YIP) under Award Number N00014-24-1-2098. R. M. was supported in part by the Simons Investigator grant (990660) and by the Simons Collaboration on Ultra-Quantum Matter, which is a grant from the Simons Foundation (651442). R. M. also acknowledges partial support from grant NSF PHY-2309135 and the Simons Investigator program (C. Xu).

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: PROOF OF LEMMA 1

Recall the variational definition of the trace norm:

$$\|X\|_1 = \sup_{\|O\| \leq 1} \text{Tr}(XO). \quad (\text{A1})$$

In our case $X = \rho^C - \tilde{\rho}^C$. Since both ρ^C and $\tilde{\rho}^C$ belong to the algebra $\mathcal{A}_C$, it suffices to consider O chosen from $\mathcal{A}_C$.

Assuming that V is supported on A and defining $\rho(s)$ as the Gibbs state of $H(s)$, then for any unit-norm operator O supported on B_2C :

$$\begin{aligned} \text{Tr}[(\rho(0) - \rho(1))O] &= -\frac{1}{2} \int_0^1 ds \text{Cov}_{\rho(s)}(\Phi^s(V), O) \\ &\quad + \text{Cov}_{\rho(s)}(O, \Phi^s(V)), \end{aligned} \quad (\text{A2})$$

where $\text{Cov}_\rho(O_1, O_2) := \text{Tr}(\rho O_1 O_2) - \text{Tr}(\rho O_1) \text{Tr}(\rho O_2)$ denotes the connected correlation in the state ρ , while $\Phi^s(V)$ is the quantum belief-propagation operator defined as [46]

$$\Phi^s(V) := \int_{-\infty}^{\infty} dt f(t) e^{-itH(s)} V e^{itH(s)} \quad (\text{A3})$$

for some exponentially decaying L_1 function $f(t)$ [see Eq. (26) in Ref. [44] for an explicit expression].

The operator $\Phi_\beta^s(V)$ has the following properties, crucial for us:

1. $\|\Phi^s(V)\| \leq \|V\|$
2. $\Phi^s(V)$ belongs to $\mathcal{A}$, the algebra generated by local interaction terms $\mathbf{h}$.
3. $\Phi^s(V)$ can be approximated by another operator $\Phi^{s,l}(V)$ supported strictly in the l -neighborhood of A , denoted by $A_{+l} := \{x \in \Lambda | d(x, A) \leq l\}$, up to an error [44,57]

$$\|\Phi_\beta^s(V) - \Phi_\beta^{s,l}(V)\| \leq \|V\| \zeta_{\text{QBP}}(A, l), \quad (\text{A4})$$

where $\zeta_{\text{QBP}}(A, l)$ decays exponentially when $H(s)$ is a short-range Hamiltonian,

$$\zeta_{\text{QBP}}(A, l) \leq 6|A| e^{-\frac{1}{1+v_{\text{LR}}/\pi} l}. \quad (\text{A5})$$

Here v_{LR} is the maximum Lieb-Robinson velocity of the Hamiltonians $H(s)_{s \in [0,1]}$. Choosing $l < d(A, C)$, we obtain

$$\begin{aligned} |\text{Tr}[(\rho(0) - \rho(1))O]| &\leq \int_0^1 ds \text{Cov}_{\rho(s)}(\Phi^{s,l}(V), O) + \|V\| \cdot \zeta_{\text{QBP}}(A, l) \\ &\leq \|V\| \left(\sup_{s \in [0,1]} \text{Cov}_{\rho(s)}^A(A_{+l}, C) + \zeta_{\text{QBP}}(A, l) \right) \\ &\leq \|V\| \left(\sup_{s \in [0,1]} \text{Cov}_{\rho(s)}^A(A_{+l}; C) + 6|A| e^{-\frac{1}{1+v/\pi} l} \right). \end{aligned} \quad (\text{A6})$$

APPENDIX B: PROOF OF LEMMA 3

For any state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, there exists a completely positive trace-preserving map $\mathcal{R}_{B \rightarrow BC} : \mathcal{B}(\mathcal{H}_B) \mapsto \mathcal{B}(\mathcal{H}_B \otimes \mathcal{H}_C)$, such that [49,58]

$$I(A : C|B)_\rho \geq -2 \log_2 F(\rho, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})), \quad (\text{B1})$$

where $F(\cdot, \cdot)$ denotes the fidelity between two states, and

$$\mathcal{R}_{B \rightarrow BC}(\cdot) := \int_{-\infty}^{\infty} dt \beta_0(t) \mathcal{R}_{\rho^{BC}, \text{Tr}_C}^t(\cdot) \quad (\text{B2})$$

is the rotated Petz recovery map. Here, $\beta_0(t) = \frac{\pi}{2} (\cosh(\pi t) + 1)^{-1}$, and for any reference state σ and quantum channel $\mathcal{E}$, we define

$$\mathcal{R}_{\sigma, \mathcal{E}}^t(\cdot) := \sigma^{\frac{1}{2} - it} \mathcal{E}^\dagger \left[\mathcal{E}(\sigma)^{-\frac{1}{2} + it} (\cdot) \mathcal{E}(\sigma)^{-\frac{1}{2} - it} \right] \sigma^{\frac{1}{2} + it}. \quad (\text{B3})$$

Furthermore, using the Fuchs-van de Graaf inequality,

$$1 - F(\rho, \sigma) \geq \frac{1}{4} \|\rho - \sigma\|_1^2, \quad (\text{B4})$$

we deduce that

$$I(A: C|B)_\rho \geq \frac{1}{4 \ln 2} \|\rho - \mathcal{R}_{B \rightarrow BC}(\rho^{AB})\|_1^2. \quad (\text{B5})$$

APPENDIX C: SYMMETRIC LOCAL INTERACTION VARIATION: EXAMPLE OF ISING SYMMETRY

In this appendix we discuss conditions under which local interaction variation can be made to respect global symmetries present in local interaction terms $\mathbf{h}$. We focus on the special case where $G = \{\mathbb{I}, g = \prod_i X_i\}$ is the global Ising symmetry. We assume each h_X commutes with g , i.e., $[h_X, g] = 0$ for all X . This implies any Gibbs state ρ_β is weakly symmetric under the Ising symmetry, i.e., $g\rho_\beta g^{-1} = \rho_\beta$.

As we recall, local Hamiltonian variation Eq. (13) is achieved by composing Tr_{AB_1} and a recovery map. It is mathematically equivalent to replace the partial trace with a complete depolarizing channel, defined as $\mathcal{D}_{AB_1}[\rho] = 2^{-|AB_1|} \text{Tr}_{AB_1}[\rho] \otimes \mathbb{I}_{AB_1}$. Therefore $\mathcal{M}$ can be rewritten as

$$\mathcal{M}[\rho_\beta] = \mathcal{R}_{B_2 \rightarrow AB} \circ \mathcal{D}_{AB_1}[\rho_\beta]. \quad (\text{C1})$$

The recovery map $\mathcal{R}_{B_2 \rightarrow AB}$ is strongly G -symmetric, as shown below:

$$\begin{aligned} g\rho_\beta^X g^{-1} &= g_X \rho_\beta^X g_X^{-1} = g_X \text{Tr}_X[g_X \rho_\beta g_X^{-1}] g_X^{-1} = \text{Tr}_X[g\rho_\beta g^{-1}] \\ &= \rho_\beta^X. \end{aligned} \quad (\text{C2})$$

In the following we discuss when the complete depolarization channel $\mathcal{D}_{AB_1}$'s action on ρ_β can be simulated by a strongly symmetric channel.

We consider performing complete depolarization in different symmetry sectors separately, as detailed below. For any region X , the state ρ_β can be decomposed according to the eigenvalue of $g_X := \prod_{i \in X} X_i$ upon both left and right actions:

$$\rho_\beta = \sum_{s_1, s_2 \in \{1, -1\}} \rho_{s_1 s_2}, \quad (\text{C3})$$

where each $\rho_{s_1 s_2} = P_{s_1} \rho P_{s_2}$ with P_s being the projector to the subspace of g_X with eigenvalue s . This can equivalently be understood as the block form of ρ_β in terms of the ± 1 eigenspaces of g_X . The symmetrized depolarization channel is defined as

$$\begin{aligned} \mathcal{D}_X^{\text{sym}}[\rho] &= \mathbb{E}_{U_+, U_-} [(U_+ \oplus U_-) \rho (U_+ \oplus U_-)^\dagger] \\ &= 2^{-|X|+1} (\text{Tr}_X(\rho_{++}) \otimes P_+ + \text{Tr}_X(\rho_{--}) \otimes P_-), \end{aligned} \quad (\text{C4})$$

where U_+ and U_- are Haar random unitaries supported on the ± 1 eigenspaces of g_X , respectively. In comparison, the complete depolarization channel's action on ρ_β is given by:

$$\mathcal{D}_X[\rho_\beta] = 2^{-|X|} \text{Tr}_X(\rho_{++} + \rho_{--}) \otimes (P_+ + P_-). \quad (\text{C5})$$

The trace distance between the two states is given by (noting $g_X = P_+ - P_-$):

$$\begin{aligned} \|\mathcal{D}_X^{\text{sym}}[\rho_\beta] - \mathcal{D}_X[\rho_\beta]\|_1 &= 2^{-|X|} \|\text{Tr}_X(\rho_{++}) \otimes g_X - \text{Tr}_X(\rho_{--}) \otimes g_X\|_1 \\ &= \|\text{Tr}_X(\rho_{++}) - \text{Tr}_X(\rho_{--})\|_1 \\ &= \|\text{Tr}_X(\rho_\beta P_+) - \text{Tr}_X(\rho_\beta P_-)\|_1 \\ &= \|\text{Tr}_X(g_X \rho_\beta)\|_1. \end{aligned} \quad (\text{C6})$$

Physically, the expectation value of the symmetry generator restricted to X can be interpreted as a disorder parameter for the *strong* Ising symmetry. Because the Gibbs state ρ_β lacks a strong Ising symmetry from the outset, this disorder parameter is expected to decay exponentially with the volume $|X|$ —in fact, for classical Gibbs states it is exactly zero. Consequently, we expect our local interaction-variation channel can be arranged to respect an on-site Ising symmetry, up to errors that are exponentially small in the volume of AB_1 in Fig. 3.

APPENDIX D: DERIVATION OF EQ. (22)

We introduce the shorthand notation $\mathcal{G}_{\geq t} = \mathcal{M}^{(N)} \circ \dots \circ \mathcal{M}^{(t)}$. We then have

$$\begin{aligned} &\|\mathcal{M}^{(N)} \circ \dots \circ \mathcal{M}^{(1)}(\rho_\beta) - \rho_{\beta+\Delta\beta}\|_1 \\ &\leq \left\| \sum_{i=0}^{N-1} (\mathcal{G}_{\geq(i+1)}[\rho^{(i+1)}] - \mathcal{G}_{\geq i}[\rho^{(i)}]) \right\|_1 \\ &\leq \sum_{i=0}^{N-1} \|\mathcal{G}_{\geq(i+1)}[\rho^{(i+1)}] - \mathcal{G}_{\geq i}[\rho^{(i)}]\|_1 \\ &\leq \sum_{i=0}^{N-1} \|\rho^{(i+1)} - \mathcal{M}^{(i)}[\rho^{(i)}]\|_1, \end{aligned} \quad (\text{D1})$$

where the last inequality uses the monotonicity of the trace norm under the action of a quantum channel.

APPENDIX E: PROOF OF LOCAL REVERSIBILITY

Let $\mathcal{M}^{(m)}$ be a gate in the t th layer of the constructed circuit $\mathcal{C}$. One can find an index $n \leq m$, such that $\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(n)}$ are gates in the first $(t-1)$ layers, and gates $\mathcal{M}^{(n+1)}, \dots, \mathcal{M}^{(m-1)}$ are in the same layer as $\mathcal{M}^{(m)}$ and are in front of $\mathcal{M}^{(m)}$ in the prescribed ordering.

We use $\mathcal{C}' := \mathcal{M}^{(n)} \circ \dots \circ \mathcal{M}^{(1)}$ to denote the first $(t-1)$ layers of the circuit. Similarly to Eq. (21), we denote $\rho^{(n)} := \rho_{\beta+\Delta\beta_n}$. Then we have

$$\begin{aligned} & \|\widetilde{\mathcal{M}}^{(m)} \circ \mathcal{M}^{(m)} \circ \mathcal{C}'[\rho_\beta] - \mathcal{C}'[\rho_\beta]\|_1 \\ & \leq 2\|\mathcal{C}'[\rho_\beta] - \rho^{(n)}\|_1 + \|\widetilde{\mathcal{M}}^{(m)} \circ \mathcal{M}^{(m)}[\rho^{(n)}] - \rho^{(n)}\|_1 \\ & \leq 2\epsilon_C, \end{aligned} \quad (\text{E1})$$

which establishes the local reversibility of $\mathcal{C}$ with respect to ρ_β .

APPENDIX F: GAPPED GROUND STATES

Reference [30] showed that for a gapped Hamiltonian whose excitations are point-like quasiparticles, $\rho_{s\beta}$ approximates the ground-state density matrix $\rho_{\text{g.s.}}$ within an error ϵ when $s \geq \log(L/\epsilon)/\Delta$. Here we demonstrate that the same result holds when the excitations reside on higher-dimensional closed manifolds (e.g., loops or membranes) with an energy cost proportional to their volume. We expect this assumption on the excitation spectrum to cover a broad class of physically relevant systems.

Consider a Hamiltonian on a D -dimensional regular lattice whose excitations live on closed p -dimensional surfaces with $p < D$. In the excitation basis, the Gibbs state reads

$$\rho_{s\beta} = \frac{1}{Z} \sum_{\{\gamma\}} e^{-sE_\gamma} |\gamma\rangle\langle\gamma|, \quad (\text{F1})$$

where γ labels an excitation configuration. As shown in Ref. [25], the probability that $\rho_{s\beta}$ is supported on configurations containing an excitation of volume V is bounded by

$$P(V) \leq \Omega_V e^{-s\Delta V}, \quad (\text{F2})$$

where Ω_V is the number of closed p -dimensional surfaces of volume V and Δ is the energy cost per unit volume. On a regular lattice one has $\Omega_V \leq L^D (c_p)^V$, with c_p the coordination constant (e.g., $D-p$ for a hypercubic lattice). Consequently, the total probability of non-ground-state configurations satisfies

$$\sum_{V>0} P(V) \leq \frac{c_p L^D e^{-s\Delta}}{1 - c_p e^{-s\Delta}}. \quad (\text{F3})$$

Since the Gibbs state is diagonal in the energy basis, we have

$$\|\rho_{\text{g.s.}} - \rho_{s\beta}\|_1 = 2 \sum_{V>0} P(V). \quad (\text{F4})$$

It follows that choosing $s = \mathcal{O}(\ln(L/\epsilon))$ ensures this trace distance is at most ϵ .

APPENDIX G: COMPARISON OF 2D AND 4D TORIC CODES

As a simple example, we illustrate in this section that the 2D toric code does not satisfy stable clustering as we approach $T = 0$. To do so, it is sufficient to find a correlation function in the annular-shaped partition that does not decay as in Eq. (3).

Let us consider the square lattice in Fig. 5, where qubits reside on the links. The 2D toric code Hamiltonian is

$$H = -\sum_{\square} A_{\square} - \sum_{+} B_{+}, \quad (\text{G1})$$

where $A_{\square} = \prod_{i \in \square} Z_i$ and $B_{+} = \prod_{i \in +} X_i$ are the usual plaquette and vertex terms, respectively. Define two loop operators supported on the loops γ_1 and γ_2 (the red and blue loops in Fig. 5, respectively):

$$O_1 = \prod_{i \in \gamma_1} Z_i, \quad O_2 = \prod_{i \in \gamma_2} X_i. \quad (\text{G2})$$

Let us denote the region enclosed by γ_1 as A , and that by γ_2 as B . We begin with a Gibbs state ρ_β at uniform inverse temperature β_0 , such that ρ_β approximates the ground-space state $\rho_{\text{g.s.}}$ within error ϵ . As discussed in Sec. F, it suffices to choose $\beta_0 \geq \log(\text{poly } L/\epsilon)$. We then choose $\Delta\beta$ so that in the state $\rho_{\beta+\Delta\beta}$, all Hamiltonian terms fully supported within region A are at a finite inverse temperature β , while those supported nontrivially outside γ_1 remain unchanged.

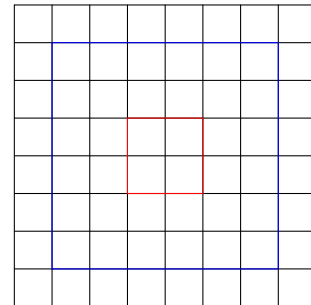


FIG. 5. Illustration of loop operators in 2D toric code.

We now compute the connected correlation function of the two closed-loop operators O_1 and O_2 in the state $\rho_{\beta+\Delta\beta}$. The operators O_1 and O_2 simply measure the parity of m anyons (the plaquette excitations) within regions A and B , respectively:

$$\begin{aligned}\text{Tr}(\rho_{\beta+\Delta\beta} O_1) &= \prod_{i \in A} \text{Tr}(\rho_{\beta+\Delta\beta} P_i) = \left(\frac{1 - e^{-2\beta}}{1 + e^{-2\beta}} \right)^{N_A} = (\tanh \beta)^{N_A}, \\ \text{Tr}(\rho_{\beta+\Delta\beta} O_2) &= \prod_{i \in B} \text{Tr}(\rho_{\beta+\Delta\beta} P_i) = (\tanh \beta)^{N_A} (\tanh \beta_0)^{N_B - N_A},\end{aligned}\tag{G3}$$

where P_i measures the parity of the i th plaquette, and N_A is the total number of plaquettes in A , and similarly for N_B . On the other hand, $O_1 O_2$ measures the parity of the m anyons within $B \setminus A$:

$$\text{Tr}(\rho_{\beta+\Delta\beta} O_1 O_2) = (\tanh \beta_0)^{N_B - N_A},\tag{G4}$$

For the state ρ_β to approximate the ground space within ϵ , we must have

$$(\tanh \beta_0)^{N_B - N_A} = \text{Tr}(\rho_{\beta+\Delta\beta} O_1 O_2) = \text{Tr}(\rho_\beta O_1 O_2) > 1 - \epsilon.\tag{G5}$$

Therefore,

$$\text{Tr}(\rho_{\beta+\Delta\beta} O_1 O_2) - \text{Tr}(\rho_{\beta+\Delta\beta} O_1) \text{Tr}(\rho_{\beta+\Delta\beta} O_2) > (1 - \epsilon)[1 - (\tanh \beta)^{2N_A}] > 0.\tag{G6}$$

This nonzero connected correlation demonstrates that (δ, ξ) -stable clustering fails in the 2D toric code when the δ -ball in parameter space attempts to approach the zero-temperature limit from a finite temperature.

In contrast, in the 4D toric code the analogous connected correlation vanishes. On a 4D hypercubic lattice, the Hamiltonian is

$$H_{4D} = -\sum_s A_s - \sum_c B_c,\tag{G7}$$

where qubits reside on two-dimensional plaquettes, and

$$A_s = \prod_{i \in \partial s} Z_i, \quad B_c = \prod_{i \ni c} X_i\tag{G8}$$

are the stabilizer terms associated with three-dimensional cubes (denoted by $\{s\}$) and one-dimensional edges (denoted by $\{c\}$), respectively. In the annular partition, the region A is the four-dimensional solid ball B^4 . We then consider the state $\rho_{\beta+\Delta\beta}$ in which all stabilizers acting nontrivially outside A are in the zero-temperature limit, while those supported entirely within A remain at finite inverse temperature β .

Let $\bar{A} = \Lambda \setminus A \simeq S^3$ denote the complement of region A on the lattice. As in the main text, let $\mathcal{A}$ be the algebra generated by all Hamiltonian terms, and define $\mathcal{A}_{\bar{A}} = \text{Tr}_A \mathcal{A}$. Moreover, let $\mathcal{B}_{\bar{A}}$ be the subalgebra generated by Hamiltonian terms supported entirely in $\bar{A}$. Evidently, $\mathcal{B}_{\bar{A}} \subset \mathcal{A}_{\bar{A}}$. Compared with the 2D case, the 4D case has a crucial distinction: In the case where $\bar{A}$ is isomorphic to S^3 ,

one can verify that $\mathcal{B}_{\bar{A}} = \mathcal{A}_{\bar{A}}$, i.e., any product of stabilizers acting nontrivially in A cannot yield an operator whose support lies entirely in $\bar{A}$ that is *not* itself a product of stabilizers contained in $\bar{A}$ [59]. Consequently, to examine the covariance between A and $\bar{A}$ as defined in Eq. (2), it suffices to consider the correlation

$$\text{Tr}(\rho_{\beta+\Delta\beta} O_1 O_2) - \text{Tr}(\rho_{\beta+\Delta\beta} O_1) \text{Tr}(\rho_{\beta+\Delta\beta} O_2),\tag{G9}$$

where $O_1 \in \mathcal{A}_A$ and $O_2 \in \mathcal{A}_{\bar{A}} = \mathcal{B}_{\bar{A}}$.

Now, by taking the temperature of all terms acting nontrivially outside A to zero, the state $\rho_{\beta+\Delta\beta}$ expels all excitations from $\bar{A}$. More precisely, if β_0 is chosen such that $\|\rho_\beta - \rho_{\text{g.s.}}\|_1 \leq \epsilon$, then by the same argument as in Sec. F, all stabilizers supported in $\bar{A}$ have eigenvalue 1 up to an error ϵ , so

$$\begin{aligned}\text{Tr}(\rho_{\beta+\Delta\beta} O_1 O_2) &\stackrel{\epsilon}{=} \text{Tr}(\rho_{\beta+\Delta\beta} O_1) \\ &\stackrel{\epsilon}{=} \text{Tr}(\rho_{\beta+\Delta\beta} O_1) \text{Tr}(\rho_{\beta+\Delta\beta} O_2),\end{aligned}\tag{G10}$$

where $O_1 \in \mathcal{A}_A$ and $O_2 \in \mathcal{B}_{\bar{A}}$, both of operator norm 1, indicating that the annular-shaped covariance vanishes in the state $\rho_{\beta+\Delta\beta}$.

We emphasize that the latter part does not provide a direct proof that stable clustering holds for the 4D toric code in the low-temperature phase; rather, it offers physical understanding into the differences between the 2D and 4D cases.

APPENDIX H: LOCAL INDISTINGUISHABILITY OF QUANTUM CODES

Suppose Π is a code space projector of a diameter- ℓ quantum code, with codespace dimension 2^k . We introduce a reference system R and have it maximally entangled with the code space, such that the total state is

$$|\Psi_{RQ}\rangle = 2^{-k/2} \sum_{a=1}^{2^k} |a\rangle_R \otimes |\tilde{a}\rangle_Q, \quad (\text{H1})$$

where $\{|a\rangle\}$ is a basis of R , and $\{|\tilde{a}\rangle\}$ is a basis of the code space Π . We have used Q to denote the physical qubits supporting the code space Π . Let $A \subset Q$ be a region with diameter less than ℓ . Then, by the definition of topological codes, the code states $\{|\tilde{a}\rangle_Q\}$ are indistinguishable on A , leading to

$$I_{|\Psi\rangle}(R:A) = 0. \quad (\text{H2})$$

Following the theorem in Ref. [60], the Hilbert space $\mathcal{H}_{\tilde{A}}$ admits a decomposition $\mathcal{H}_{\tilde{A}} = (\mathcal{H}_X \otimes \mathcal{H}_Y) \otimes \mathcal{H}'$ under which the state $|\Psi_{RQ}\rangle$ takes the form

$$|\Psi_{RQ}\rangle = 2^{-k/2} \sum_{a=1}^{2^k} |a\rangle_R \otimes |\tilde{a}'\rangle_{RX} \otimes |\omega_0\rangle_{YA}, \quad (\text{H3})$$

such that $|\omega_0\rangle_{YA}$ is a fixed state on $Y \cup A$ independent of a . Now, for any two code states ρ_Q and σ_Q within the codespace Π , under the decomposition above, they must be of the form

$$\rho_Q = \rho'_X \otimes |\omega_0\rangle\langle\omega_0|_{YA} \quad \sigma_Q = \sigma'_X \otimes |\omega_0\rangle\langle\omega_0|_{YA}. \quad (\text{H4})$$

Letting $\mathcal{D}[\cdot] = \text{Tr}_X(\cdot)\rho'_X$ and $\mathcal{D}'[\cdot] = \text{Tr}_X(\cdot)\sigma'_X$, we have $\mathcal{D}[\sigma] = \rho$ and $\mathcal{D}'[\rho] = \sigma$. Since $\mathcal{D}$ and $\mathcal{D}'$ are local operations supported on $\bar{A}$, ρ and σ are $(\ell, 0)$ -LI.

APPENDIX I: LOCAL REVERSIBILITY AND LOCAL INDISTINGUISHABILITY

A few notations for this appendix: $\rho \stackrel{\delta}{=} \sigma$ means $\|\rho - \sigma\|_1 \leq \delta$. $\mathcal{G} = \prod_X \mathcal{E}_X$ is used to denote one layer of a channel circuit. The gate $\mathcal{E}_X$ acts within a simply connected region X . Different X 's do not overlap. $\text{rng}(\mathcal{G}) := \sup_X \text{diam}(X) \cdot |\mathcal{G}|$ means the number of gates in $\mathcal{G}$.

Def: A layer $\mathcal{G} = \prod_X \mathcal{E}_X$ is locally reversible (LR) with respect to ρ if, for each X , there exists a channel gate $\tilde{\mathcal{E}}_X$ supported within X such that $\tilde{\mathcal{E}}_X \circ \mathcal{E}_X[\rho] \stackrel{\delta}{=} \rho$.

Claim: If $\mathcal{G}$ is δ -LR with respect to ρ , then $\tilde{\mathcal{G}} \circ \mathcal{G}[\rho] \stackrel{|\mathcal{G}| \cdot \delta}{=} \rho$. The proof is straightforward.

Def: ρ and σ are (r, δ) locally indistinguishable (LI) if, for any simply connected A with $\text{diam}(A) < r$, there exist channels $\mathcal{D}, \mathcal{D}'$ supported within $\bar{A}$ such that $\mathcal{D}[\rho] \stackrel{\delta}{=} \sigma, \mathcal{D}'[\sigma] \stackrel{\delta}{=} \rho$.

For instance, any two codeword states σ, ρ of the toric code on an $L \times L$ torus are $(L/2, 0)$ -LI.

Claim: If ρ and σ are (r, δ_{LI}) -LI, and $\mathcal{G}$ is δ_{LR} -LR with respect to σ , then:

1. $\mathcal{G}$ is $(2\delta_{LI} + \delta_{LR})$ -L.R. with respect to ρ .
2. $\mathcal{G}[\rho]$ and $\mathcal{G}[\sigma]$ are $(r - 2\text{rng } \mathcal{G}, (2|\mathcal{G}| + 1)\delta_{LI} + |\mathcal{G}|\delta_{LR})$ -LI

Proof of the first claim is straightforward. For each X ,

$$\tilde{\mathcal{E}}_X \circ \mathcal{E}_X[\rho] \stackrel{\delta_{LI}}{=} \tilde{\mathcal{E}}_X \circ \mathcal{E}_X \circ \mathcal{D}[\sigma] = \mathcal{D} \circ \tilde{\mathcal{E}}_X \circ \mathcal{E}_X[\sigma] \stackrel{\delta_{LR}}{=} \mathcal{D}[\sigma] \stackrel{\delta_{LI}}{=} \rho \quad (\text{I1})$$

where $\mathcal{D}$ is chosen to be within $\bar{X}$.

To prove the second claim, let A be any region with $\text{diam}(A) \leq r - 2\text{rng}(\mathcal{G})$. Let B be a width- $\text{rng } \mathcal{G}$ buffer around A , and let $C = \bar{A}B$. Since $\text{diam}(AB) \leq r$, there exists $\mathcal{D}$ in C such that $\mathcal{D}[\rho] \stackrel{\delta_{LI}}{=} \sigma$.

Let $\mathcal{G} = \mathcal{G}_C \mathcal{G}_{\bar{C}}$ where $\mathcal{G}_C$ contains all gates having overlap with C , and $\mathcal{G}_{\bar{C}}$ contains the remaining gates. Then $\mathcal{G}_{\bar{C}}$ is supported within AB . Let $\mathcal{D}^{\mathcal{G}} := \mathcal{G}_C \mathcal{D} \tilde{\mathcal{G}}_C$ and $\mathcal{D}'^{\mathcal{G}} := \mathcal{G}_C \mathcal{D}' \tilde{\mathcal{G}}_C$, both acting on BC . Then

$$\begin{aligned} \mathcal{D}^{\mathcal{G}} \circ \mathcal{G}[\sigma] &= \mathcal{G}_C \mathcal{D}' \tilde{\mathcal{G}}_C \mathcal{G}_C \mathcal{G}_{\bar{C}}[\sigma] \stackrel{|\mathcal{G}_C| \cdot \delta_{LR}}{=} \mathcal{G}_C \mathcal{D}' \mathcal{G}_{\bar{C}}[\sigma] \stackrel{\delta_{LI}}{=} \mathcal{G}[\rho] \\ \mathcal{D}'^{\mathcal{G}} \circ \mathcal{G}[\rho] &= \mathcal{G}_C \mathcal{D} \tilde{\mathcal{G}}_C \mathcal{G}_C \mathcal{G}_{\bar{C}}[\rho] \stackrel{|\mathcal{G}_C| \cdot (2\delta_{LI} + \delta_{LR})}{=} \mathcal{G}_C \mathcal{D} \mathcal{G}_{\bar{C}}[\rho] \stackrel{\delta_{LI}}{=} \mathcal{G}[\sigma], \end{aligned} \quad (\text{I2})$$

which completes the proof.

We now generalize the result to circuits with multiple layers. Consider a circuit $\mathcal{C} = \mathcal{G}_T \mathcal{G}_{T-1} \dots \mathcal{G}_1$ such that each layer $\mathcal{G}_t$ is δ -LR with respect to $\mathcal{C}_{t-1}[\sigma]$, where $\mathcal{C}_t := \mathcal{G}_t \mathcal{G}_{t-1} \dots \mathcal{G}_1$ (i.e., the first t layers of $\mathcal{C}$). Further, let ρ and σ be $(r, 0)$ -LI. Consider any region A with $\text{diam}(A) \leq r - 2\text{rng } \mathcal{C}$, and let B be a width- $\text{rng } \mathcal{C}$ buffer region that shields A from $C = \bar{A}B$. Analogously, define channels $\mathcal{D}_t := \mathcal{C}_{t,C} \mathcal{D} \tilde{\mathcal{C}}_{t,C}$ and $\mathcal{D}'_t := \mathcal{C}_{t,C} \mathcal{D}' \tilde{\mathcal{C}}_{t,C}$ ($t = 0, \dots, T$), where $\mathcal{C}_{t,C}$ contains all gates in $\mathcal{C}_t$ that overlap with the *light cone* of region C , with $\tilde{\mathcal{C}}_{t,C}$ defined analogously so that it consists of the local inverses of the gates in $\mathcal{C}_{t,C}$. Since $\mathcal{D}$ and $\mathcal{D}'$ are supported in region C , for $t \leq T$, $\mathcal{D}_t$ and $\mathcal{D}'_t$ have support outside region A . Since $\mathcal{C}$ is δ -LR with respect to σ , it follows that for $t \leq T$,

$$\mathcal{D}'_t \circ \mathcal{C}_t[\sigma] \stackrel{|\mathcal{C}_{t,C}| \cdot \delta}{=} \mathcal{C}_t \circ \mathcal{D}'[\sigma] = \mathcal{C}_t[\rho]. \quad (\text{I3})$$

As a result, a gate $\mathcal{E}_{t,X}$ in $\mathcal{G}_t$ that is supported on A is $(2|\mathcal{C}_{t-1,C}| + 1)\delta$ -LR with respect to $\mathcal{C}_{t-1}[\rho]$:

$$\tilde{\mathcal{E}}_{t,X} \circ \mathcal{E}_{t,X} \circ \mathcal{C}_{t-1}[\rho] \stackrel{|C_{t-1,C}|}{=} \tilde{\mathcal{E}}_{t,X} \circ \mathcal{E}_{t,X} \circ \mathcal{D}'_{t-1} \circ \mathcal{C}_{t-1}[\sigma] \stackrel{\delta}{=} \mathcal{D}'_{t-1} \circ \mathcal{C}_{t-1}[\sigma] \stackrel{|C_{t-1,C}|}{=} \mathcal{C}_{t-1}[\rho]. \quad (I4)$$

Consequently,

$$\mathcal{D}_T \circ \mathcal{C}_T[\rho] \stackrel{\epsilon}{=} \mathcal{C}_T[\sigma] \quad \epsilon = |\mathcal{C}_{T,C}| \cdot (2|\mathcal{C}_{T,C}| + 1), \quad (I5)$$

which implies that $\mathcal{C}[\rho]$ and $\mathcal{C}[\sigma]$ are $(r - 2\text{rng } \mathcal{C}, 2|\mathcal{C}|^2 + |\mathcal{C}|)$ -LI. This completes the proof of Lemma 4.

APPENDIX J: DERIVATION OF EQ. (31)

Here we prove the following claim: if a Lindbladian $\mathcal{L}$ and a state ρ satisfy $\|\mathcal{L}[\rho]\|_1 \leq \epsilon$, then $\|e^{t\mathcal{L}}[\rho] - \rho\|_1 \leq t\epsilon$. For any τ we have

$$\begin{aligned} \|e^{\tau\mathcal{L}}(\rho) - \rho\|_1 &= \left\| \sum_{n=1}^{\infty} \frac{\tau^n}{n!} \mathcal{L}^n[\mathcal{L}(\rho)] \right\|_1 \\ &\leq \sum_{n=1}^{\infty} \frac{\tau^n}{n!} \|\mathcal{L}^n[\mathcal{L}(\rho)]\|_1 \\ &\leq \sum_{n=1}^{\infty} \frac{\tau^n}{n!} \|\mathcal{L}\|_1^{n-1} \cdot \epsilon \\ &\leq \sum_{n=1}^{\infty} \frac{\tau^n}{n!} \|\mathcal{L}\|_1^n \cdot \frac{\epsilon}{\|\mathcal{L}\|_1} \\ &= (e^{\tau\|\mathcal{L}\|_1} - 1) \frac{\epsilon}{\|\mathcal{L}\|_1} \end{aligned} \quad (J1)$$

Suppose $t = m \cdot \tau$ with m being an integer. Then we have

$$\begin{aligned} \|e^{t\mathcal{L}}(\rho) - \rho\|_1 &\leq \sum_{k=0}^{m-1} \|e^{(k+1)\tau\mathcal{L}}[\rho] - e^{k\tau\mathcal{L}}[\rho]\|_1 \\ &\leq \sum_{k=0}^{m-1} \|e^{\tau\mathcal{L}}[\rho] - \rho\|_1 \\ &\leq (e^{t\|\mathcal{L}\|_1/m} - 1) \frac{m \cdot \epsilon}{\|\mathcal{L}\|_1} \\ \lim_{m \rightarrow \infty} &= \epsilon \cdot t \end{aligned} \quad (J2)$$

-
- [1] J. McGreevy, *Generalized symmetries in condensed matter*, *Annu. Rev. Condens. Matter Phys.* **14**, 57 (2023).
 - [2] X.-G. Wen, *Colloquium: Zoo of quantum-topological phases of matter*, *Rev. Mod. Phys.* **89**, 041004 (2017).
 - [3] R. M. Nandkishore and M. Hermele, *Fractons*, *Annu. Rev. Condens. Matter Phys.* **10**, 295 (2019).
 - [4] X. Chen, Z.-C. Gu, and X.-G. Wen, *Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order*, *Phys. Rev. B* **82**, 155138 (2010).

- [5] J. Preskill, *Quantum computing in the NISQ era and beyond*, *Quantum* **2**, 79 (2018).
- [6] A. Coser and D. Pérez-García, *Classification of phases for mixed states via fast dissipative evolution*, *Quantum* **3**, 174 (2019).
- [7] R. Ma and C. Wang, *Average symmetry-protected topological phases*, *Phys. Rev. X* **13**, 031016 (2023).
- [8] M. B. Hastings, *Topological order at nonzero temperature*, *Phys. Rev. Lett.* **107**, 210501 (2011).
- [9] T. Rakovszky, S. Gopalakrishnan, and C. von Keyserlingk, *Defining stable phases of open quantum systems*, *Phys. Rev. X* **14**, 041031 (2024).
- [10] R. Sohal and A. Prem, *Noisy approach to intrinsically mixed-state topological order*, *PRX Quantum* **6**, 010313 (2025).
- [11] T. D. Ellison and M. Cheng, *Toward a classification of mixed-state topological orders in two dimensions*, *PRX Quantum* **6**, 010315 (2025).
- [12] Y.-H. Chen and T. Grover, *Separability transitions in topological states induced by local decoherence*, *Phys. Rev. Lett.* **132**, 170602 (2024).
- [13] Y.-H. Chen and T. Grover, *Symmetry-enforced many-body separability transitions*, *PRX Quantum* **5**, 030310 (2024).
- [14] C. de Groot, A. Turzillo, and N. Schuch, *Symmetry protected topological order in open quantum systems*, *Quantum* **6**, 856 (2022).
- [15] R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, *Topological phases with average symmetries: The decohered, the disordered, and the intrinsic*, *Phys. Rev. X* **15**, 021062 (2025).
- [16] Z. Wang, Z. Wu, and Z. Wang, *Intrinsic mixed-state topological order*, *PRX Quantum* **6**, 010314 (2025).
- [17] L. A. Lessa, R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, *Strong-to-weak spontaneous symmetry breaking in mixed quantum states*, *PRX Quantum* **6**, 010344 (2025).
- [18] L. A. Lessa, S. Sang, T.-C. Lu, T. H. Hsieh, and C. Wang, *Higher-form anomaly and long-range entanglement of mixed states*, [arXiv:2503.12792](https://arxiv.org/abs/2503.12792).
- [19] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, *Mixed-state topological order and the errorfield double formulation of decoherence-induced transitions*, [arXiv:2301.05687](https://arxiv.org/abs/2301.05687).
- [20] R. Fan, Y. Bao, E. Altman, and A. Vishwanath, *Diagnostics of mixed-state topological order and breakdown of quantum memory*, *PRX Quantum* **5**, 020343 (2024).
- [21] J. Y. Lee, C.-M. Jian, and C. Xu, *Quantum criticality under decoherence or weak measurement*, *PRX Quantum* **4**, 030317 (2023).
- [22] T.-C. Lu, Z. Zhang, S. Vijay, and T. H. Hsieh, *Mixed-state long-range order and criticality from measurement and feedback*, *PRX Quantum* **4**, 030318 (2023).

- [23] C. Zhang, Y. Xu, J.-H. Zhang, C. Xu, Z. Bi, and Z.-X. Luo, *Strong-to-weak spontaneous breaking of 1-form symmetry and intrinsically mixed topological order*, *Phys. Rev. B* **111**, 115137 (2025).
- [24] T.-H. Yang, B. Shi, and J. Y. Lee, *Topological mixed states: Axiomatic approaches and phases of matter*, *arXiv:2506.04221*.
- [25] S.-T. Zhou, M. Cheng, T. Rakovszky, C. von Keyserlingk, and T. D. Ellison, *Finite-temperature quantum topological order in three dimensions*, *arXiv:2503.02928*.
- [26] P. Sala, J. Alicea, and R. Verresen, *Decoherence and wavefunction deformation of $D=4$ non-Abelian topological order*, *Phys. Rev. X* **15**, 031002 (2025).
- [27] L. A. Lessa, M. Cheng, and C. Wang, *Mixed-state quantum anomaly and multipartite entanglement*, *Phys. Rev. X* **15**, 011069 (2025).
- [28] M. B. Hastings and X.-G. Wen, *Quasiadiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance*, *Phys. Rev. B* **72**, 045141 (2005).
- [29] M. B. Hastings, *Locality in quantum systems*, *arXiv:1008.5137*.
- [30] M. B. Hastings, *Entropy and entanglement in quantum ground states*, *Phys. Rev. B* **76**, 035114 (2007).
- [31] Models displaying discrete symmetry-breaking order, i.e., the Ising model, satisfy our notion of clustering since all symmetric local observables have rapidly decaying correlation functions.
- [32] S. Sang, L. A. Lessa, R. S. K. Mong, T. Grover, C. Wang, and T. H. Hsieh, *Mixed-state phases from local reversibility*, *arXiv:2507.02292*.
- [33] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Topological quantum memory*, *J. Math. Phys.* **43**, 4452 (2002).
- [34] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, *On thermal stability of topological qubit in Kitaev's 4D model*, *arXiv:0811.0033*.
- [35] C.-F. Chen and C. Rouzé, *Quantum Gibbs states are locally Markovian*, *arXiv:2504.02208*.
- [36] R. Gheissari and A. Sinclair, *Low-temperature Ising dynamics with random initializations*, *arXiv:2106.11296*.
- [37] T. Bergamaschi, R. Gheissari, and Y. Liu, *Rapid mixing for Gibbs states within a logical sector: A dynamical view of self-correcting quantum memories*, *arXiv:2507.10976*.
- [38] M. Hasenbusch, S. Meyer, and M. Pütz, *The roughening transition of the three-dimensional Ising interface: A Monte Carlo study*, *J. Stat. Phys.* **85**, 383 (1996).
- [39] T. P. Eggarter, *Cayley trees, the Ising problem, and the thermodynamic limit*, *Phys. Rev. B* **9**, 2989 (1974).
- [40] A. Montanari and G. Semerjian, *Rigorous inequalities between length and time scales in glassy systems*, *J. Stat. Phys.* **125**, 23 (2006).
- [41] Y. Hong, J. Guo, and A. Lucas, *Quantum memory at nonzero temperature in a thermodynamically trivial system*, *Nat. Commun.* **16**, 316 (2025).
- [42] B. Placke, G. M. Sommers, N. P. Breuckmann, T. Rakovszky, and V. Khemani, *Expansion creates spin-glass order in finite-connectivity models: A rigorous and intuitive approach from the theory of LDPC codes*, *arXiv:2507.13342*.
- [43] B. Placke, T. Rakovszky, N. P. Breuckmann, and V. Khemani, *Topological quantum spin glass order and its realization in QLDPC codes*, *arXiv:2412.13248*.
- [44] Á. Capel, M. Moscolari, S. Teufel, and T. Wessel, *From decay of correlations to locality and stability of the Gibbs state*, *Commun. Math. Phys.* **406**, 43 (2025).
- [45] R. König and F. Pastawski, *Generating topological order: No speedup by dissipation*, *Phys. Rev. B* **90**, 045101 (2014).
- [46] M. B. Hastings, *Quantum belief propagation: An algorithm for thermal quantum systems*, *Phys. Rev. B* **76**, 201102 (2007).
- [47] M. S. Leifer and D. Poulin, *Quantum graphical models and belief propagation*, *Ann. Phys. (Amsterdam)* **323**, 1899 (2008).
- [48] W. Brown and D. Poulin, *Quantum Markov networks and commuting Hamiltonians*, *arXiv:1206.0755*.
- [49] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, *Universal recovery maps and approximate sufficiency of quantum relative entropy*, *arXiv:1509.07127*.
- [50] Following standard definition, ground-state subspace includes all states whose energy splitting from the ground state is smaller than any $\text{Poly}^{-1}L$.
- [51] We consider quasi-local rather than strictly local Lindbladians since, for any Gibbs state ρ_β of local Hamiltonians, there exists a quasi-local Lindbladian $\mathcal{L} = \sum_X \mathcal{L}_X$ having ρ_β as a frustration-free steady state, i.e., $\mathcal{L}_X[\rho_\beta] = 0$ for each X [52]. It is unclear whether strictly local Lindbladians with the same property exist for generic noncommuting Hamiltonians.
- [52] C.-F. Chen, M. J. Kastoryano, and A. Gilyén, *An efficient and exact noncommutative quantum Gibbs sampler*, *arXiv:2311.09207*.
- [53] We emphasize that a code's diameter is not its distance: A quantum code has distance d if $\prod O \prod \propto \prod$ for any operator O supported on a region with number of sites no more than d .
- [54] The converse does not hold in general: Two states can agree on all local observables without being related by a channel acting outside A . Actually, as shown in Ref. [32], matching of local observables, combined with the approximate Markov property, is equivalent to Definition 4.
- [55] G. Kalachev and S. Sadov, *A linear-algebraic and lattice-theoretical look at the cleaning lemma of quantum coding theory*, *Linear Algebra Appl.* **649**, 96 (2022).
- [56] A. Harrow, S. Mehraban, and M. Soleimanifar, *Classical algorithms, correlation decay, and complex zeros of partition functions of quantum many-body systems*, *arXiv:1910.09071*.
- [57] Á. M. Alhambra, *Quantum many-body systems in thermal equilibrium*, *PRX Quantum* **4**, 040201 (2023).
- [58] O. Fawzi and R. Renner, *Quantum conditional mutual information and approximate Markov chains*, *Commun. Math. Phys.* **340**, 575 (2015).
- [59] In 2D, the operator O_2 in Eq. (G2) lies in $A_{\bar{A}}$ but not in $B_{\bar{A}}$.
- [60] P. Hayden, R. Jozsa, D. Petz, and A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, *Commun. Math. Phys.* **246**, 359 (2004).