Separability Transitions in Topological States Induced by Local Decoherence

Yu-Hsueh Chen and Tarun Grover
Department of Physics, University of California at San Diego, La Jolla, California 92093, USA

(Received 12 November 2023; accepted 1 April 2024; published 26 April 2024)

We study states with intrinsic topological order subjected to local decoherence from the perspective of *separability*, i.e., whether a decohered mixed state can be expressed as an ensemble of short-range entangled pure states. We focus on toric codes and the X-cube fracton state and provide evidence for the existence of decoherence-induced separability transitions that precisely coincide with the threshold for the feasibility of active error correction. A key insight is that local decoherence acting on the "parent" cluster states of these models results in a Gibbs state. As an example, for the 2D (3D) toric code subjected to bit-flip errors, we show that the decohered density matrix can be written as a convex sum of short-range entangled states for $p > p_c$, where p_c is related to the paramagnetic-ferromagnetic transition in the 2D (3D) random bond Ising model along the Nishimori line.

DOI: 10.1103/PhysRevLett.132.170602

In this Letter, we will explore aspects of many-body topological states subjected to decoherence from the perspective of *separability*, i.e., whether the resulting mixed state can be expressed as a convex sum of short-range entangled (SRE) states [1–3]. This criteria is central to the definition of what constitutes an SRE or long-range entangled (LRE) mixed state, and various measures of mixed-state entanglement, such as negativity [3–8] and entanglement of formation [9], are defined so as to quantify nonseparability. We will be particularly interested in decoherence-induced "separability transitions," i.e., transitions tuned by decoherence such that the density matrix in one regime is expressible as a convex sum of SRE states, and in the other regime, it is not. One salient distinction between pure-state versus mixed-state dynamics is that although a short-depth unitary evolution cannot change long-range entanglement encoded in a pure state, a short-depth local *channel* can fundamentally alter long-range mixed-state entanglement. Therefore, even the limited class of mixed states that are obtained by the action of local short-depth channels on an entangled pure state offer an opportunity to explore mixed-state phases and phase transitions [10–22]. We will focus on mixed states that are obtained via subjecting several well-understood topologically ordered phases of matter to short-depth quantum channels.

Error-threshold theorems [23–28] suggest a topologically ordered pure state is perturbatively stable against decoherence from a short-depth, local quantum channel, leading to the possibility of a phase transition as a function of the decoherence rate [29]. Such transitions were originally studied from the perspective of quantum error correction (QEC) in Refs. [30,31] and more recently using mixed-state entanglement measures such as topological negativity [14], and other nonlinear functions of the density matrix (Refs. [13–15]). These approaches clearly establish at least two different mixed-state phases: one where the topological qubit can be decoded, and the other where it cannot. However, it is not obvious if the density matrix in the regime where decoding fails can be expressed as a convex sum of SRE pure states, which, following Refs. [1,2], we will take as the definition of an SRE mixed state. Our main result is that for several topologically ordered phases subjected to local decoherence, which are relevant for quantum computing [30–32], one can explicitly write down the decohered mixed state as a convex sum of pure states which we argue all undergo a topological phase transition, from being long-ranged entangled to short-ranged entangled, at a threshold that precisely corresponds to the optimal threshold for QEC. We find that the universality class of such a separability transition also coincides with that corresponding to the QEC error-recovery transition. Therefore, in these examples, we argue that the error-recovery transition does indeed coincide with a many-body separability transition. As discussed below, our method also provides a new route to obtain the statistical mechanics models relevant for the quantum error-correcting codes [14,30,31,33].

Let us begin by considering the ground state of the 2D toric code [see Fig. 1(b)] with Hamiltonian $H_{\mathrm{2D\,toric}} = -\sum_v(\prod_{e\in v} Z_e) - \sum_p(\prod_{e\in p} X_e)$ subjected to phase-flip errors. The Hilbert space consists of qubits residing on the edges (denoted as "e") of a square lattice and we assume periodic boundary conditions. Denoting the ground state as ρ_0, the Kraus map corresponding to the phase-flip errors act on an edge e as $\mathcal{E}_e[\rho_0] = pZ_e\rho_0 Z_e + (1-p)\rho_0$, and the full map is given by the composition of this map over all edges. The key first step is to utilize the idea of duality [34–38] by identifying the corresponding "parent" cluster Hamiltonian

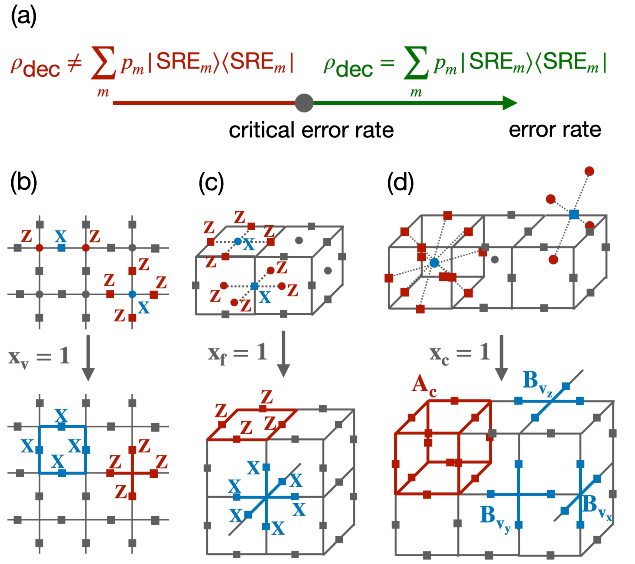

FIG. 1. (a) Topological orders under local decoherence can undergo a separability transition, where only above a certain critical error rate, the decohered mixed state $\rho_{\rm dec}$ can be written as a convex sum of SRE pure states. The bottom depicts the parent cluster states and their offspring models obtained by appropriate measurements (indicated by an arrow); (b) 2D cluster Hamiltonian and 2D toric code; (c) 3D cluster Hamiltonian and 3D toric code; and (d) "Cluster-X" Hamiltonian [39] and the X-cube Hamiltonian.

(in the sense of Refs. [39–42]). Interestingly, the application of the aforementioned Kraus map to its ground state results in a Gibbs state. For the problem at hand, consider $H_{\rm 2D\,cluster} = \sum_v h_v + \sum_e h_e$, where $h_v = -X_v(\prod_{e\ni v} Z_e)$ and $h_e = -X_e(\prod_{v\in e} Z_v)$ whose Hilbert space consists of qubits both on the vertices and the edges of the square lattice [Fig. 1(b)]. The ground state density matrix ρ_0 of the 2D toric code can be written as $\rho_0 \propto \langle x_{\mathbf{v}} = 1|\rho_{C,0}|x_{\mathbf{v}} = 1\rangle$, where $|x_{\mathbf{v}} = 1\rangle = \otimes_v |x_v = 1\rangle$ is the product state in the Pauli-X basis, and $\rho_{C,0}[\propto \prod_e(I-h_e)\prod_v(I-h_v)]$ is the ground state of the $H_{\rm 2D\,cluster}$. The projection selects one specific ground state of the toric code that is an eigenvector of the noncontractible Wilson loops $W_\ell = \prod_{e\in\ell} X_e$ with eigenvalue $+1$ along both cycles ℓ of the torus. A simple calculation shows that $\mathcal{E}_e[\rho_{C,0}] \propto e^{-\beta\sum_e h_e}\prod_v(I-h_v)$ where $\tanh(\beta) = 1\text{–}2p$. This implies that the decohered density matrix ρ of the toric code is $\rho \propto \langle x_{\mathbf{v}} = 1|e^{-\beta\sum_e h_e}|x_{\mathbf{v}} = 1\rangle P_Z$, where $P_Z = \prod_v(I + \prod_{e\ni v} Z_e)$. By inserting a complete set of states, one may simplify the above expression as $\rho \propto P_Z\rho_e P_Z$ where $\rho_e = \sum_{x_{\mathbf{e}}} \mathcal{Z}_{{\rm 2D\,Ising},x_{\mathbf{e}}}|x_{\mathbf{e}}\rangle\langle x_{\mathbf{e}}|$ and $\mathcal{Z}_{{\rm 2D\,Ising},x_{\mathbf{e}}} = \sum_{z_{\mathbf{v}}} e^{\beta\sum_e x_e\prod_{v\in e} z_v}$ is the partition function of the 2D Ising model with Ising interactions determined by $\{x_e\}$. Thus, $\rho \propto \sum_{x_{\mathbf{e}}} \mathcal{Z}_{{\rm 2D\,Ising},x_{\mathbf{e}}}|\Omega_{x_{\mathbf{e}}}\rangle\langle\Omega_{x_{\mathbf{e}}}|$, where $|\Omega_{x_{\mathbf{e}}}\rangle \propto \prod_v(I + \prod_{e\ni v} Z_e)|x_{\mathbf{e}}\rangle$ are nothing but a subset of toric code eigenstates. Note that in this derivation, the 2D Ising model emerges due to the h_e terms in the parent cluster Hamiltonian, and ultimately, this will lead to the relation between the separability transition and the statistical mechanics of the 2D random-bond Ising model (RBIM) that also describes the error-recovery transition [30]. We note that the above spectral representation of ρ in terms of toric code eigenstates has also previously appeared in Ref. [13], using a different derivation. Since noncontractible cycles of the torus will play an important role below, let us note that distinct eigenstates $|\Omega_{x_{\mathbf{e}}}\rangle$ can be uniquely specified by two labels: the first label corresponds to the set of local $\mathbb{Z}_2$ fluxes $f_p = \prod_{e\in p} x_e$ through elementary plaquettes p, while the second label $\mathbf{L} = (L_x = \pm 1, L_y = \pm 1)$ with $L_x = \prod_{e\in\ell, e\parallel\hat{x}} x_e$, $L_y = \prod_{e\in\ell, e\parallel\hat{y}} x_e$ and ℓ a noncontractible loop along $\hat{x}/\hat{y}$ direction, specifies the topological sector ($\mathbf{L}$ representing "logical data") in which $|\Omega_{x_{\mathbf{e}}}\rangle$ lives.

We now probe the mixed state ρ using the separability criteria, i.e., we ask whether it can be decomposed as a convex sum of SRE states. Clearly, the aforementioned spectral representation is not a useful decomposition since it involves toric code eigenstates which are all LRE. Taking a cue from the argument for separability of the Gibbs state of toric codes [43], we decompose ρ as $\rho = \sum_{z_{\mathbf{e}}} \rho^{1/2}|z_{\mathbf{e}}\rangle\langle z_{\mathbf{e}}|\rho^{1/2} \equiv \sum_m |\psi_m\rangle\langle\psi_m|$ where $\{z_{\mathbf{e}}\}$ are a complete set of product states in the Pauli-Z basis, and $|\psi_m\rangle = \rho^{1/2}|z_{\mathbf{e}}\rangle$. Generically, to determine whether ρ is an SRE mixed state, one needs to determine whether *each* $|\psi_m\rangle$ is SRE. However, for the current case of interest, it suffices to consider only $|\psi\rangle = \rho^{1/2}|m_0\rangle$ with $|m_0\rangle = |z_{\mathbf{e}} = 1\rangle$. The reason is as follows. The Gauss's law $(\prod_{e\ni v} Z_e = 1)$ implies that the Hilbert space only contains states that are closed loops in the Z basis. Therefore, one may write $|m\rangle = g_x|m_0\rangle$ where g_x is a product of *single-site* Pauli-Xs forming closed loops. Since $[g_x, \rho] = 0$, this implies that $|\psi_m\rangle \equiv |\psi_{g_x}\rangle = g_x|\psi\rangle$, and therefore, if $|\psi\rangle$ is SRE (LRE), so is $|\psi_{g_x}\rangle$. $\rho(\beta)$ may then be written as $\rho(\beta) = \sum_{g_x}|\psi_{g_x}(\beta)\rangle\langle\psi_{g_x}(\beta)|$. Now, using the aforementioned spectral representation of ρ, the (non-normalized) state $|\psi\rangle = \rho^{1/2}|z_{\mathbf{e}} = 1\rangle$ is

$$|\psi(\beta)\rangle \propto \sum_{x_{\mathbf{e}}}[\mathcal{Z}_{{\rm 2D\,Ising},x_{\mathbf{e}}}(\beta)]^{1/2}|x_{\mathbf{e}}\rangle. \quad (1)$$

It is easy to see that when $\beta = \infty$, $|\psi\rangle \propto |\Omega_0\rangle$, the non-decohered toric code ground state, while when $\beta = 0$, $|\psi\rangle \propto |z_{\mathbf{e}} = 1\rangle$ is a product state. This suggests a phase transition for $|\psi(\beta)\rangle$ from being an LRE state to an SRE state as we increase the error rate p (i.e., decrease β). We will now show that this is indeed the case.

We first consider the expectation value of the "anyon condensation operator" (also known as 't Hooft loop) in state $|\psi(\beta)\rangle$, defined as [15,44–46] $T_{\tilde{\ell}} = \prod_{e\in\tilde{\ell}} Z_e$, where $\tilde{\ell}$ denotes a homologically noncontractible loop on the dual lattice (in the language of $\mathbb{Z}_2$ gauge theory [46,47],

$Z_e \sim e^{i\pi(\text{electric field})_e}$). Physically, $\langle T_{\tilde{\ell}}\rangle \equiv \langle\psi|T_{\tilde{\ell}}|\psi\rangle/\langle\psi|\psi\rangle$ is the amplitude of tunneling from one logical subspace to an orthogonal one, and therefore it is zero in the $\mathbb{Z}_2$ topologically ordered phase, and nonzero in the topologically trivial phase (= anyon condensed phase) [48]. Indeed, one may easily verify that $\langle T_{\tilde{\ell}}\rangle = 0(1)$ when $\beta = \infty(\beta = 0)$. Using Eq. (1), $T_{\tilde{\ell}}$ flips spins along the curve $\tilde{\ell}$ (i.e., $x_e \to -x_e,\ \forall\ e \in \tilde{\ell}$), and we denote the corresponding configuration as $x_{\tilde{\ell},\mathbf{e}}$. While $x_{\tilde{\ell},\mathbf{e}}$ and $x_\mathbf{e}$ have the same flux through every elementary plaquette, they live in different logical sectors $\mathbf{L}$. Therefore, $T_{\tilde{\ell}}|\psi\rangle \propto \sum_{x_\mathbf{e}}[\mathcal{Z}_{x_\mathbf{e}}]^{1/2}|x_{\tilde{\ell},\mathbf{e}}\rangle = \sum_{x_\mathbf{e}}[\mathcal{Z}_{x_{\tilde{\ell},\mathbf{e}}}]^{1/2}|x_\mathbf{e}\rangle$, where we have suppressed the subscript "2D Ising" under the partition function $\mathcal{Z}$ for notational convenience. Thus,

$$\langle T_{\tilde{\ell}}\rangle = \frac{\sum_{x_\mathbf{e}}\sqrt{\mathcal{Z}_{x_\mathbf{e}}\mathcal{Z}_{x_{\tilde{\ell},\mathbf{e}}}}}{\sum_{x_\mathbf{e}}\mathcal{Z}_{x_\mathbf{e}}} = \frac{\sum_{x_\mathbf{e}}\mathcal{Z}_{x_\mathbf{e}}e^{-\Delta F_{x_{\tilde{\ell},\mathbf{e}}}/2}}{\sum_{x_\mathbf{e}}\mathcal{Z}_{x_\mathbf{e}}} = \langle e^{-\Delta F_{\tilde{\ell}}/2}\rangle \geq e^{-\langle\Delta F_{\tilde{\ell}}\rangle/2}, \tag{2}$$

where $\Delta F_{x_{\tilde{\ell},\mathbf{e}}} = -\log(\mathcal{Z}_{x_{\tilde{\ell},\mathbf{e}}}/\mathcal{Z}_{x_\mathbf{e}})$ is the free energy cost of inserting a domain wall of size $|\tilde{\ell}| \sim L$ (= system's linear size) in the RBIM along the Nishimori line [30], and we have used Jensen's inequality in the last sentence. We are along the Nishimori line because the probability of a given gauge invariant label $\{f_\mathbf{p}, \mathbf{L}\}$ along the Nishimori line is precisely the partition function $\mathcal{Z}_{x_\mathbf{e}}$ [30]. Since $\langle\Delta F_{\tilde{\ell}}\rangle$, the disorder-averaged free energy cost, diverges with L in the ferromagnetic phase of the RBIM while converges to a constant in the paramagnetic phase [30], Equation (2) rigorously shows that for $p > p_c = p_{\text{2D RBIM}} \approx 0.109$ [49], $\langle T_{\tilde{\ell}}\rangle$ saturates to a nonzero constant. Therefore, $|\psi\rangle$ is a topologically trivial state when $p > p_c$, and hence the mixed state is SRE for $p > p_c$. In contrast, for $p < p_c$, due to nonvanishing ferromagnetic order (and associated domain wall cost) of the RBIM, we expect that $\langle T_{\tilde{\ell}}\rangle \sim e^{-\langle\Delta F_{\tilde{\ell}}\rangle/2} \sim e^{-cL} \to 0$ in the thermodynamic limit ($c > 0$ is a constant), implying that $|\psi\rangle$ is topologically ordered. This does not necessarily imply that the decohered state ρ is long-range entangled for $p < p_c$, because there may exist some other way to decompose it as a sum of SRE pure states. However, for $p < p_c$, long-range entanglement as quantified by topological entanglement negativity is nonzero as shown in Ref. [14]. Since a state of a form $\rho = \sum_i p_i|\text{SRE}\rangle_{ii}\langle\text{SRE}|$ can be prepared by an ensemble of finite-depth unitary circuits starting with a product state, it is reasonable to assume that it cannot support long-range entanglement as quantified by any valid measure of entanglement. Assuming that topological negativity is one such measure (as supported by previous works [14,43]), the above discussion implies that the state of our interest undergoes a separability transition at $p = p_c$.

Another diagnostic of topological order in pure states is the (Renyi) topological entanglement entropy (TEE) [50–52]. Dividing the system in real space as $A \cup B$, and defining the reduced density matrix $\rho_A = \text{tr}_B|\psi(\beta)\rangle\langle\psi(\beta)|$ for the state $|\psi(\beta)\rangle$ [Eq. (1)], one finds [53]:

$$\text{tr}(\rho_A^2) = \frac{\sum_{x_\mathbf{e},x'_\mathbf{e}}\mathcal{Z}_{x_A,x_B}\mathcal{Z}_{x'_A,x'_B}e^{-\Delta F_{AB}(x_\mathbf{e},x'_\mathbf{e})/2}}{\sum_{x_\mathbf{e},x'_\mathbf{e}}\mathcal{Z}_{x_A,x_B}\mathcal{Z}_{x'_A,x'_B}}. \tag{3}$$

Here, $x_A(x_B)$ denotes all the edges belonging to the region $A(B)$, $\mathcal{Z}_{x_A,x_B}$ denotes the partition function of the 2D Ising model with the sign of Ising interactions determined by x_A and x_B, and $\Delta F_{AB}(x_\mathbf{e}, x'_\mathbf{e}) = -\log[\mathcal{Z}_{x'_A,x_B}\mathcal{Z}_{x_A,x'_B}/(\mathcal{Z}_{x_A,x_B}\mathcal{Z}_{x'_A,x'_B})]$ is the free energy cost of swapping bonds between two copies of RBIM in region A. We provide a heuristic argument (Ref. [53]) that the TEE jumps from log (2) to zero at $p_c = p_{\text{2D RBIM}}$. The main idea is that in the ferromagnetic phase of the RBIM, the free energy penalty of creating a single Ising vortex leads to a specific nonlocal constraint on the allowed configurations that contribute to the sum in Eq. (3). The constraint is essentially that one needs to minimize the free energy cost $\Delta F_{AB}(x_\mathbf{e}, x'_\mathbf{e})$ for each fixed flux configuration $\{f_p\}$ and $\{f'_p\}$ corresponding to $\{x_e\}$ and $\{x'_e\}$ in Eq. (3). One finds that there always exists a *pair* of configurations that contribute equally to $\text{tr}(\rho_A^2)$ while satisfying the aforementioned constraint. This results in a subleading contribution of $-\log(2)$ in the entanglement entropy, which we identify as TEE. In the paramagnetic phase, the aforementioned nonlocal constraint does not exist, and one therefore does not expect a nonzero TEE. Therefore, we arrive at the same conclusion as the one obtained using the anyon condensation operator.

Incidentally, one may also construct an alternative convex decomposition of the decohered mixed state ρ that shows a phase transition at a certain (nonoptimal) threshold $p_{\text{nonoptimal}}$ which is related to 2D RBIM via a Kramers-Wannier duality [34]. The main outcome is that $\tanh(\beta_{\text{nonoptimal}}) = 1\text{–}2p_{\text{nonoptimal}}$ satisfies $\tanh^2(\beta_{\text{nonoptimal}}/2) = p_{\text{2D RBIM}}/(1 - p_{\text{2D RBIM}})$ which yields $p_{\text{nonoptimal}} \approx 0.188$. See Ref. [53] for details.

Let us next consider the 3D toric code with $H_{\text{3D toric}} = -\sum_f(\prod_{e\in f}Z_e) - \sum_v(\prod_{e\in v}X_e)$ [see Fig. 1(c)] subjected to phase-flip errors (local nontrivial Kraus operators $\sim Z_e$). Previous work has already identified error recovery transition at $p_c \approx 0.029$ with universality determined by the 3D random-plaquette gauge model (RPGM) along the Nishimori line [31]. We first verify that the corresponding mixed-state density matrix is long-range entangled for $p < p_c$ by calculating its entanglement negativity. We find a nonzero, quantized topological negativity log(2) for $p < p_c$ and zero for $p > p_c$. The calculation is conceptually similar to that for the 2D toric code [14], see Ref. [53] for details. Having confirmed the presence of long-range entanglement for $p < p_c$, we now ask whether the mixed-state at $p > p_c$ is separable? To proceed, we follow a strategy similar to that for the 2D toric code, and rewrite the ground state of 3D toric code in terms of the ground state $\rho_{C,0}$ of a parent cluster state with Hamiltonian $H_{\text{3D cluster}} = -\sum_e X_e \prod_{f\ni e} Z_f - \sum_f X_f \prod_{e\in f} Z_e = \sum_e h_e + \sum_f h_f$. The corresponding ground state density matrix of the 3D toric code ρ_0 on a three-torus is $\rho_0 \propto \langle x_{\mathbf{f}} = 1|\rho_{C,0}|x_{\mathbf{f}} = 1\rangle$, which is an eigenstate of the noncontractible 't Hooft membrane operators T_{xy}, T_{yz}, T_{zx} along the three planes with eigenvalue $+1$ ($T_{xy} = \prod_{e\|z} X_e$ where the product is taken over all edges parallel to the z axis in any xy plane; T_{yz} and T_{zx} are defined analogously). Following essentially the same steps as in 2D toric code, one obtains the decohered density matrix $\rho = \sum_{g_x} |\psi_{g_x}(\beta)\rangle\langle\psi_{g_x}(\beta)|$ with $|\psi_{g_x}\rangle = g_x|\psi(\beta)\rangle$ and g_x a product of single-site Pauli-Xs forming closed membranes. Therefore, we again only need to analyze whether $|\psi\rangle \equiv \rho^{1/2}|z_{\mathbf{e}} = 1\rangle$ is SRE or LRE. Again, one may rewrite $|\psi(\beta)\rangle \propto \sum_{x_{\mathbf{e}}} [\mathcal{Z}_{\text{3D gauge},x_{\mathbf{e}}}(\beta)]^{1/2}|x_{\mathbf{e}}\rangle$ where $\mathcal{Z}_{\text{3D gauge},x_{\mathbf{e}}} = \sum_{z_{\mathbf{f}}} e^{\beta\sum_e x_e \prod_{f\ni e} z_f}$ is now the partition function of a classical 3D Ising gauge theory with the sign of each plaquette term determined by $\{x_e\}$. To probe the topological transition in $|\psi\rangle$ as a function of β, we now consider the Wilson loop operator $W_\ell = \prod_{e\in\ell} Z_e$, where ℓ denotes a homologically nontrivial cycle on the original lattice, say, along the z axis (so that it pierces and anticommutes with T_{xy}). One finds $\langle W_\ell\rangle = \langle e^{-\Delta F_\ell/2}\rangle \geq e^{-\langle\Delta F_\ell\rangle/2}$, where ΔF_ℓ now denotes the free energy cost of inserting a domain wall along the noncontractible loop for the 3D RPGM along the Nishimori line. Since $\langle\Delta F_\ell\rangle$ diverges as the length of $|\mathcal{C}| \sim L$ (= system size) in the Higgs (ordered) phase, while it converges to a constant in the confinement (disordered) phase [31], one finds $\langle W_\ell\rangle$ saturates to a nonzero constant when $p > p_{\text{3D RPGM}} \approx 0.029$ [31], while it vanishes for $p < p_c$. Therefore, $|\psi\rangle$, and correspondingly the decohered state ρ, is SRE when $p > p_{\text{3D RPGM}}$ and LRE for $p < p_{\text{3D RPGM}}$. One can similarly study the 3D toric code with bit-flip errors. In this case, one finds that the separability transition is dictated by the transition out of the ferromagnetic phase in the 3D RBIM along the Nishimori line, which matches the optimal error-recovery threshold, $p_c \approx 0.233$ [54,55]. See Ref. [53] for details.

Finally, let us briefly consider the 3D X-cube model (Ref. [61]), where the Hilbert space consists of qubits residing on the edges (e) of a cubic lattice, and the Hamiltonian is $H_{\text{X cube}} = -\sum_c \prod_{e\in c} Z_e - \sum_v (\prod_{e\in v_x} X_e + \prod_{e\in v_y} X_e + \prod_{e\in v_z} X_e) = -\sum_c A_c - \sum_v (B_{v_x} + B_{v_y} + B_{v_z})$ where $e \in v_\gamma, \gamma = x, y, z$ denotes all the edges emanating from the vertex v that are normal to the γ direction [see Fig. 1(d)]. Previous work has already established that under local decoherence, this system undergoes an error-recovery transition at $p_c \approx 0.152$ with universality determined by 3D plaquette Ising model [62]. We now show that for $p > p_c$, the density matrix can be written as a convex sum of SRE pure states. We exploit the observation in Ref. [39] that the ground state density matrix ρ_0 of the X-cube model can be written as $\rho_0 \propto \langle x_{\mathbf{c}} = 1|\rho_{C,0}|x_{\mathbf{c}} = 1\rangle$, where $\rho_{C,0} \propto \prod_c (I - h_c) \prod_e (I - h_e)$ ($h_c = -X_c \prod_{e\in c} Z_e$ and $h_e = -X_e \prod_{c\ni e} Z_c$) denotes the ground state density matrix of the parent cluster state, and $|x_{\mathbf{c}} = 1\rangle = \otimes_c |x_c = 1\rangle$ is the product state in the Pauli-X basis. The qubits in the parent cluster state live at the edges and the centers of the cubes so that h_c involves 13-qubit interactions, and h_e involves 5-qubit interactions [Fig. 1(d)]. The density matrix after subjecting ρ_0 to the phase-flip channel (Kraus operators $\sim Z_e$) can be written as $\rho \propto \sum_{x_{\mathbf{e}}} \mathcal{Z}_{\text{3D plaquette},x_{\mathbf{e}}} |\Omega_{x_{\mathbf{e}}}\rangle\langle\Omega_{x_{\mathbf{e}}}|$, where $|\Omega_{x_{\mathbf{e}}}\rangle \propto \prod_e (I + \prod_{e\in c} Z_e)|x_{\mathbf{e}}\rangle$ and $\mathcal{Z}_{\text{3D plaquette},x_{\mathbf{e}}} = \sum_{z_{\mathbf{c}}} e^{\beta\sum_e x_e \prod_{c\ni e} z_c}$ is the partition function of the 3D plaquette Ising model [63] with the sign of interaction on each plaquette determined by $\{x_e\}$. One again only needs to analyze the state $|\psi\rangle = \rho^{1/2}|z_{\mathbf{e}} = 1\rangle$ to study the separability transition for ρ. Now there exist exponentially many topological sectors [61], and in the nondecohered ground state ρ_0, the membrane operators defined as $\prod_{e\|\hat{a}} X_e$ with $a = x, y, z$ for *any* plane have expectation value one. To detect the presence or absence of topological order in $|\psi\rangle$, one therefore considers noncontractible Wilson loop operators $W_\ell = \prod_{e\in\ell} Z_e$ that anticommute with the membrane operators orthogonal to ℓ. The expectation value of any such Wilson loop takes a form similar to Eq. (2) where the partition function $\mathcal{Z}_{x_{\mathbf{e}}} = \mathcal{Z}_{\text{3D plaquette},x_{\mathbf{e}}}$ and one is again along the Nishimori line. This again indicates that the pure state $|\psi\rangle$ undergoes a transition at the error threshold $p_c = p_{\text{3D plaquette}} \approx 0.152$ [62].

Finally, an alternative, heuristic approach to any of the phase transitions discussed above is via considering a more general class of wave functions $|\psi^{(\alpha)}\rangle \propto \rho^{\alpha/2}|z_{\mathbf{e}} = 1\rangle$, which capture the separability transition for the density matrix $\rho^{(\alpha)} \equiv \rho^\alpha/\text{tr}(\rho^\alpha)$. For example, it is known that when ρ is the decohered 2D toric code state, $|\psi^{(2)}\rangle$ undergoes a phase transition from being topologically ordered to an SRE state at a threshold $p_c^{(2)}$ that is related to the critical temperature of the 2D *translationally invariant* classical Ising model [56], and correspondingly, we find that the topological negativity of $\rho^{(2)}$ undergoes a transition from $\log 2$ to 0 at $p_c^{(2)}$ (see Ref. [53]). This transition can also be located by the wave function overlap $\log(\langle\psi^{(2)}(\beta)|\psi^{(2)}(\beta)\rangle)$, which is proportional to the free energy of the classical 2D Ising model. This motivates a generalization of this overlap to general α for any of the models considered above by defining $F_\alpha(\beta) = [1/(1-\alpha)]\log(\langle\psi^{(\alpha)}(\beta)|\psi^{(\alpha)}(\beta)\rangle)$. Taking the limit $\alpha \to 1$, which corresponds to the wave function of our main interest [Eq. (1) for the 2D toric code, and analogous states for the other two models], one finds that $F_1(\beta)$ precisely corresponds to the free energy of the corresponding statistical mechanics model along the

Nishimori line, which indeed shows a singularity at the optimal error-recovery threshold p_c.

To summarize, we showed that decoherence-induced separability transitions in several topological states coincide with the optimal threshold for QEC [13–15,30,62]. Therefore in these models, the inability to correct logical errors implies an ability to prepare the mixed state using an ensemble of short-depth unitary circuits, which is our main result. The parent cluster Hamiltonian approach we discuss also provides an alternative method to find the relevant statistical mechanics models. The convex decomposition discussed captures the universal aspects of the phase diagram, as well as the threshold correctly, and it is optimal in this sense. It will be interesting to consider generalization of our approach to non-CSS and/or non-Abelian topological states. It might also be interesting to explore connections between decohered mixed states and perturbed *pure* topological states since they are both connected to finite temperature classical phase transitions [59]. Finally, we remark that using standard duality arguments [13,35], the separability transitions discussed here may also be reformulated as transitions for quantum magnets with a global $\mathbb{Z}_2$ symmetry. For example, our result implies that the ground state of the 2D quantum Ising model with Hamiltonian $H = -\sum_{\langle i,j\rangle} \tau_i^z \tau_j^z - h \sum_i \tau_i^x$ will undergo a separability transition in the paramagnetic phase when subjected to decoherence with Kraus operators $K_{ij} \propto \tau_i^z \tau_j^z$ at strength p. Above a certain p_c (which, at $h = 0$, equals $p_{\mathrm{2D\,RBIM}} \approx 0.109$), one will enter a new phase where the density matrix is expressible as a convex sum of states where each of them spontaneously breaks the $\mathbb{Z}_2$ symmetry, while such a representation is not possible below p_c [recall that under Wegner duality, SRE states map to Greenberger-Horne-Zeilinger states]. We leave a detailed exploration of this dual description for a future study.

The authors thank Dan Arovas, Tim Hsieh, and John McGreevy for helpful discussions, and Yimu Bao for helpful comments on the manuscript. T. G. is supported by the National Science Foundation under Grant No. DMR-1752417. This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958.

[1] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[2] M. B. Hastings, Phys. Rev. Lett. **107**, 210501 (2011).
[3] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[4] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
[5] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).
[6] J. Eisert and M. B. Plenio, J. Mod. Opt. **46**, 145 (1999).
[7] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).
[8] M. B. Plenio, Phys. Rev. Lett. **95**, 090503 (2005).
[9] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[10] C. de Groot, A. Turz illo, and N. Schuch, Quantum **6**, 856 (2022).
[11] R. Ma and C. Wang, Phys. Rev. X **13**, 031016 (2023).
[12] J. Y. Lee, Y.-Z. You, and C. Xu, arXiv:2210.16323.
[13] J. Y. Lee, C.-M. Jian, and C. Xu, PRX Quantum **4**, 030317 (2023).
[14] R. Fan, Y. Bao, E. Altman, and A. Vishwanath, arXiv:2301.05689.
[15] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, arXiv:2301.05687.
[16] Y. Zou, S. Sang, and T. H. Hsieh, Phys. Rev. Lett. **130**, 250403 (2023).
[17] T.-C. Lu, Z. Zhang, S. Vijay, and T. H. Hsieh, PRX Quantum **4**, 030318 (2023).
[18] R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, arXiv:2305.16399.
[19] R. Ma, arXiv:2304.08277.
[20] K. Su, N. Myerson-Jain, and C. Xu, Phys. Rev. B **109**, 035146 (2024).
[21] Z. Wang, Z. Wu, and Z. Wang, arXiv:2307.13758.
[22] S. Sang, Y. Zou, and T. H. Hsieh, arXiv:2310.08639.
[23] P. W. Shor, in *Proceedings of the 37th Conference on Foundations of Computer Science* (1996), pp. 56–65, 10.1109/SFCS.1996.548464.
[24] D. Aharonov and M. Ben-Or, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (1997), pp. 176–188, 10.1145/258533.258579.
[25] A. Y. Kitaev, Ann. Phys. (Amsterdam) **303**, 2 (2003).
[26] E. Knill, R. Laflamme, and W. H. Zurek, Science **279**, 342 (1998).
[27] J. Preskill, Proc. R. Soc. A **454**, 385 (1998).
[28] B. M. Terhal, Rev. Mod. Phys. **87**, 307 (2015).
[29] D. Aharonov, Phys. Rev. A **62**, 062311 (2000).
[30] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys. (N.Y.) **43**, 4452 (2002).
[31] C. Wang, J. Harrington, and J. Preskill, Ann. Phys. (Amsterdam) **303**, 31 (2003).
[32] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Phys. Rev. A **86**, 032324 (2012).
[33] H. Nishimori, Prog. Theor. Phys. **66**, 1169 (1981).
[34] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[35] F. J. Wegner, J. Math. Phys. (N.Y.) **12**, 2259 (1971).
[36] D. J. Williamson, Phys. Rev. B **94**, 155128 (2016).
[37] A. Kubica and B. Yoshida, arXiv:1805.01836.
[38] T. Rakovszky and V. Khemani, arXiv:2310.16032.
[39] R. Verresen, N. Tantivasadakarn, and A. Vishwanath, arXiv:2112.03061.
[40] R. Raussendorf, S. Bravyi, and J. Harrington, Phys. Rev. A **71**, 062313 (2005).
[41] M. Aguado, G. K. Brennen, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **101**, 260501 (2008).
[42] N. Tantivasadakarn, R. Thorngren, A. Vishwanath, and R. Verresen, arXiv:2112.01519.
[43] T.-C. Lu, T. H. Hsieh, and T. Grover, Phys. Rev. Lett. **125**, 116801 (2020).
[44] G. Hooft, Nucl. Phys. **B138**, 1 (1978).
[45] J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979).

[46] E. Fradkin, *Field Theories of Condensed Matter Physics* (Cambridge University Press, Cambridge, England, 2013).

[47] J. Kogut and L. Susskind, Phys. Rev. D **11**, 395 (1975).

[48] We note that the 't Hooft loop works as an order parameter for detecting topological order in our pure state $|\psi\rangle$ because the Gauss's law ($\prod_{e \ni v} Z_e = 1$) is satisfied exactly and therefore there are no dynamical charges. Besides, note that $\mathrm{tr}(\rho T_{\tilde{\ell}})$, where ρ is the decohered mixed state, is identically zero, and clearly does not capture mixed-state entanglement.

[49] A. Honecker, M. Picco, and P. Pujol, Phys. Rev. Lett. **87**, 047201 (2001).

[50] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006).

[51] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).

[52] S. T. Flammia, A. Hamma, T. L. Hughes, and X.-G. Wen, Phys. Rev. Lett. **103**, 261601 (2009).

[53] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.170602 for details, which also includes Refs. [14,15,30,50,51,54–60].

[54] Y. Ozeki and N. Ito, J. Phys. A **31**, 5451 (1998).

[55] A. Kubica, M. E. Beverland, F. Brandão, J. Preskill, and K. M. Svore, Phys. Rev. Lett. **120**, 180501 (2018).

[56] C. Castelnovo and C. Chamon, Phys. Rev. B **77**, 054433 (2008).

[57] S. Dong, E. Fradkin, R. G. Leigh, and S. Nowling, J. High Energy Phys. 05 (2008) 016.

[58] Y. Zhang, T. Grover, A. Turner, M. Oshikawa, and A. Vishwanath, Phys. Rev. B **85**, 235151 (2012).

[59] D. J. Williamson, C. Delcamp, F. Verstraete, and N. Schuch, Phys. Rev. B **104**, 235151 (2021).

[60] J. Y. Lee, W. Ji, Z. Bi, and M. Fisher, arXiv:2208.11699.

[61] S. Vijay, J. Haah, and L. Fu, Phys. Rev. B **94**, 235157 (2016).

[62] H. Song, J. Schönmeier-Kromer, K. Liu, O. Viyuela, L. Pollet, and M. A. Martin-Delgado, Phys. Rev. Lett. **129**, 230502 (2022).

[63] G. Savvidy and F. Wegner, Nucl. Phys. **B413**, 605 (1994).