Title

Diagnostics of Mixed-State Topological Order and Breakdown of Quantum Memory

Permalink

https://escholarship.org/uc/item/6c38c58j

Journal

PRX Quantum, 5(2)

ISSN

2691-3399

Authors

Fan, Ruihua

Bao, Yimu

Altman, Ehud

et al.

Publication Date

2024-05-01

DOI

10.1103/prxquantum.5.020343

Diagnostics of Mixed-State Topological Order and Breakdown of Quantum Memory

Ruihua Fan,[1,*,†] Yimu Bao,[2,†] Ehud Altman,[2,3] and Ashvin Vishwanath[1]

[1] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[2] *Department of Physics, University of California, Berkeley, California 94720, USA*
[3] *Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 22 February 2023; revised 13 December 2023; accepted 22 December 2023; published 24 May 2024)

Topological quantum memory can protect information against local errors up to finite error thresholds. Such thresholds are usually determined based on the success of decoding algorithms rather than the intrinsic properties of the mixed states describing corrupted memories. Here we provide an intrinsic characterization of the breakdown of topological quantum memory, which both gives a bound on the performance of decoding algorithms and provides examples of topologically distinct mixed states. We employ three information-theoretical quantities that can be regarded as generalizations of the diagnostics of ground-state topological order, and serve as a definition for topological order in error-corrupted mixed states. We consider the topological contribution to entanglement negativity and two other metrics based on quantum relative entropy and coherent information. In the concrete example of the two-dimensional (2D) Toric code with local bit-flip and phase errors, we map three quantities to observables in 2D classical spin models and analytically show they all undergo a transition at the same error threshold. This threshold is an upper bound on that achieved in any decoding algorithm and is indeed saturated by that in the optimal decoding algorithm for the Toric code.

DOI: 10.1103/PRXQuantum.5.020343

I. INTRODUCTION

The major roadblock to realizing quantum computers is the presence of errors and decoherence from the environment, which can only be overcome by adopting quantum error correction (QEC) and fault tolerance [1]. A first step would be the realization of robust quantum memories [2–4]. Topologically ordered systems in two spatial dimensions, owing to their long-range entanglement and consequent degenerate ground states, serve as a promising candidate [5–8]. A paradigmatic example is the surface code [9,10], whose promise as a robust quantum memory has stimulated recent interest in its realization in near-term quantum simulators [11–17].

One of the central quests is to analyze the performance of topological quantum memory under *local* decoherence. In the case of surface code and other topological codes with local errors, it has been shown that the stored information can be decoded reliably up to a finite error threshold [10,18–22]. Namely, as the error rate increases, the success probability of the decoding algorithm drops to zero at a critical value, which depends on the choice of the algorithm. It is then natural to ask whether these decoding transitions stem from an intrinsic error-induced singularity in the mixed states. If so, how to probe this intrinsic transition?

The intrinsic characterization has at least two important consequences. First, the critical error rate for the intrinsic transition should furnish an upper bound for decoding algorithms. The algorithmic dependence of the decoding thresholds is a mere reflection of the suboptimality of specific algorithms. Second, the correspondence between successful decoding and intrinsic properties of the quantum state acted upon by errors points to the existence of topologically distinct mixed states. In another word, answering this question amounts to relating the breakdown of topological quantum memory to a transition in the mixed-state topological order. Progress towards this goal lies in quantifying the residual long-range entanglement in the error-corrupted mixed state. We will consider quantities that are motivated from both perspectives and explore their unison.

In this work, we investigate three information-theoretical diagnostics: (i) quantum relative entropy

[*] ruihuafan.phys@gmail.com
[†] These authors contributed equally to this work.

between the error-corrupted ground state and excited state; (ii) coherent information; (iii) topological entanglement negativity. The first two are natural from the perspective of quantum error correction (QEC). More specifically, the quantum relative entropy quantifies whether errors ruin orthogonality between states [23], and coherent information is known to give robust necessary and sufficient conditions on successful QEC [24–26]. The third one is a basis-independent characterization of long-range entanglement in mixed states and is more natural from the perspective of mixed-state topological order. This quantity has been proposed to diagnose topological orders in Gibbs states [27,28], which changes discontinuously at the critical temperature. We borrow and apply this proposal to error-corrupted states. Our transition occurs in two spatial dimensions at a finite error rate, in contrast to the finite temperature transitions in four spatial dimensions.

The presence of three seemingly different diagnostics raises the question of whether they all agree and share the same critical error rate. Satisfyingly, we indeed find this to be the case in a concrete example, surface code with bit-flip and phase errors. The nth Rényi version of the three quantities can be formulated in a *classical* two-dimensional statistical mechanical model of $(n-1)$-flavor Ising spins, which exhibits a transition from a paramagnetic to a ferromagnetic phase as the error rate increases. The three quantities are mapped to different probes of the ferromagnetic order and must undergo the transition simultaneously, which establishes their consistency in this concrete example.

Interestingly, the statistical mechanical model derived for the information-theoretic diagnostics is exactly dual to the random-bond Ising model (RBIM) that governs the decoding transition of the algorithm proposed in Ref. [10]. This duality implies that the error threshold of the algorithm in Ref. [10] saturates the upper bound. Therefore, it confirms that this decoding algorithm is optimal, and its threshold reflects the intrinsic properties of the corrupted state. We remark that mappings to statistical mechanical models have been tied to obtaining error thresholds of decoding algorithms [10,18–22]. Here such mappings arise from characterizing intrinsic properties of the error corrupted mixed state.

The rest of the paper is organized as follows. Section II gives a concrete definition of the error-corrupted states and introduces the three diagnostics. Section III studies the concrete example, the 2D Toric code subject to local bit-flip and phase errors. We close with discussions in Sec. IV.

II. SETUP AND DIAGNOSTICS

In this section, we begin with introducing the error-corrupted mixed state. We show that any operator expectation value in a single-copy corrupted density matrix cannot probe the transition, and instead one needs to consider the nonlinear functions of the density matrix. Next, we introduce three information-theoretic diagnostics of the transition: (i) quantum relative entropy; (ii) coherent information; (iii) topological entanglement negativity. These quantities generalize the diagnostics of ground-state topological order.

A. Error-corrupted mixed state

The type of mixed state we consider in the paper describes a topologically ordered ground state $|\Psi_0\rangle\langle\Psi_0|$ subject to local errors

$$\rho = \mathcal{N}[|\Psi_0\rangle\langle\Psi_0|] = \prod_i \mathcal{N}_i[|\Psi_0\rangle\langle\Psi_0|], \tag{1}$$

where the quantum channel $\mathcal{N}_i$ models the local error at site i and is controlled by the error rate p. We refer to ρ as the error-corrupted mixed state.

The transition in the corrupted state, if exists, cannot be probed by the operator expectation value in a single-copy density matrix. To demonstrate it, we purify the corrupted state by introducing one ancilla qubit prepared in $|0\rangle_i$ for each physical qubit at site i. The physical and ancilla qubits are coupled locally via unitary $U_i(p)$ such that tracing out the ancilla qubits reproduces the corrupted state ρ. This leads to a purification

$$|\Phi\rangle = \prod_i U_i(p)\,|\Psi_0\rangle\left(\otimes_i |0\rangle_i\right), \tag{2}$$

which is related to the topologically ordered state by a depth-1 unitary circuit on the extended system [see Fig. 1]. It follows that the expectation value of *any* operator supported on a large but finite region of the physical qubits, e.g., a Wilson loop operator, must be a smooth function of the error rate [see Fig. 1 for a schematics]. Thus, it is indispensable to consider the nonlinear functions of the density matrix, e.g., quantum information quantities, to probe the transition in the corrupted state. This property holds when ρ describes a general mixed state in the ground-state subspace under local errors.

We remark that the above argument does not prevent observables in a single-copy density matrix from detecting topological order in finite-temperature Gibbs states [29]. The key difference is the purifications of the Gibbs states at different temperatures are not necessarily related by finite-depth circuits.

B. Quantum relative entropy

Anyon excitations are crucial for storing and manipulating quantum information in a topologically ordered state. For example, to change the logical state of the code one creates a pair of anyons out of the vacuum and separates them to opposite boundaries of the system. The first

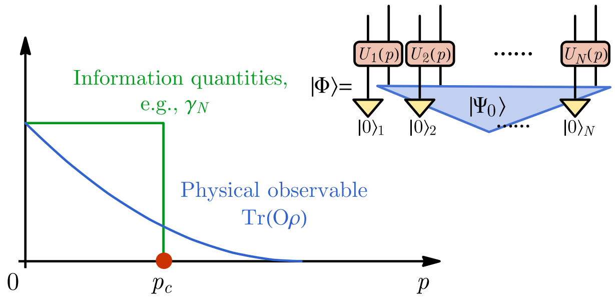


FIG. 1. Physical observables verses information quantities in error-corrupted states. Each error-corrupted state can be obtained from applying local unitaries to the system (topological order) plus ancilla qubits (trivial product state). Thus, physical observables must be smooth functions of the error rate p. In contrast, information quantities, e.g., the topological entanglement negativity γ_N, can have discontinuities that identify the many-body singularities.

diagnostic tests if the process of creating a pair of anyons and separating them by a large distance gives rise to a distinct state in the presence of decoherence.

Specifically, we want to test if the corrupted state $\rho = \mathcal{N}[\rho_0]$ is sharply distinct from $\rho_\alpha = \mathcal{N}[w_\alpha(\mathcal{P})\rho_0 w_\alpha(\mathcal{P})^\dagger]$ for ρ_0 in the ground-state subspace. In the second state, $w_\alpha(\mathcal{P})$ is an open string operator that creates an anyon α and its antiparticle α' at the opposite ends of the path $\mathcal{P}$. We use the *quantum relative entropy* as a measure for the distinguishability of the two states

$$D(\rho||\rho_\alpha) := \mathrm{tr}\rho \log \rho - \mathrm{tr}\rho \log \rho_\alpha. \tag{3}$$

In absence of errors the relative entropy is infinite because the two states are orthogonal, and it decreases monotonically with the error rate [30–32]. Below the critical error rate, however, the states should remain perfectly distinguishable if the anyons are separated by a long distance. Therefore we expect the relative entropy to diverge as the distance between the anyons is taken to infinity. Above the critical error rate on the other hand we expect the relative entropy to saturate to a finite value reflecting the inability to perfectly distinguish between the two corrupted states. In this regard, the relative entropy describes whether anyon excitations remain well defined and is a generalization of the Fredenhagen-Marcu order parameter for ground-state topological order [33–36].

To facilitate calculations, we consider a specific sequence of the Rényi relative entropies

$$D^{(n)}(\rho||\rho_\alpha) := \frac{1}{1-n}\log\frac{\mathrm{tr}\rho\rho_\alpha^{n-1}}{\mathrm{tr}\rho^n}, \tag{4}$$

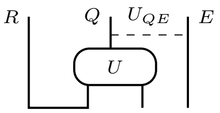
which recovers $D(\rho||\rho_\alpha)$ in the limit $n \to 1$. In Sec. III we map the relative entropies $D^{(n)}$ in the corrupted Toric code to order parameter correlation functions in an effective statistical mechanical model, which is shown to exhibit the expected behavior on two sides of the critical error rate.

C. Coherent information

The basis for protecting quantum information in topologically ordered states is encoding it in the degenerate ground state subspace. The second diagnostic we consider is designed to test the integrity of this protected quantum memory.

We use the *coherent information*, as a standard metric for the amount of recoverable quantum information after a decoherence quantum channel [24–26]. In our case, the relevant quantum channel consists of the following ingredients illustrated below: (i) a unitary operator U that encodes the state of the logical qubits in the input R into the ground-state subspace; (ii) a unitary coupling U_{QE} of the physical qubits Q to environment qubits E, which models the decoherence. The coherent information in this setup is defined as

$$I_c(R\rangle Q) := S_Q - S_{QR}. \tag{5}$$



Here S_Q and S_{RQ} are the von Neumann entropies of the systems Q and RQ, respectively, and we use the Choi map to treat the input R as a reference qubit in the output. It follows from subadditivity that the coherent information is bounded by the amount of encoded information in the degenerate ground-state subspace, i.e., $-S_R \leqslant I_c \leqslant S_R$. In the absence of errors, $I_c = S_R$, and we expect this value to persist as long as the error rate is below the critical value. Above the critical error rate, we expect $I_c < S_R$, indicating the loss of encoded information. We remark that the recoverable information is also used to characterize the robustness of quantum memory based on the edge mode in 1D Kitaev chain [37].

Physically the coherent information is closely related and expected to undergo a transition at the same point as the relative entropy discussed above. The quantum information is encoded by separating anyon pairs across the system. It stands to reason that if this state remains perfectly distinguishable from the original state, as quantified by the relative entropy, then the quantum information encoded in this process is preserved.

The critical error rate for preserving the coherent information is an upper bound for the threshold of any QEC algorithms

$$p_c \geqslant p_{c,\mathrm{algorithm}}. \tag{6}$$

The key point is that coherent information is nonincreasing upon quantum information processing and cannot be restored once it is lost. Thus, a successful QEC requires $I_c = S_R$. Moreover, the QEC algorithm involves syndrome measurements that are nonunitary and generically do not access the full coherent information in the system giving rise to a lower error threshold. To facilitate calculations and mappings to a statistical mechanical model we will need the Rényi coherent information

$$I_c^{(n)} := S_Q^{(n)} - S_{RQ}^{(n)} = \frac{1}{n-1}\log\frac{\mathrm{tr}\rho_{RQ}^n}{\mathrm{tr}\rho_Q^n}, \tag{7}$$

which approaches I_c in the limit $n \to 1$. In the example of Toric code with incoherent errors discussed in Sec. III, we show that $I_c^{(n)}$ takes distinct values in different phases.

D. Topological entanglement negativity

The topological entanglement entropy provides an intrinsic bulk probe of ground-state topological order and does not require *a priori* knowledge of the anyon excitations. The third diagnostic we consider generalizes this notion to the error-corrupted mixed state.

A natural quantity often used to quantify entanglement in mixed states, is the logarithmic negativity of a subregion A [38–40]

$$\mathcal{E}_A(\rho) := \log ||\rho^{T_A}||_1, \tag{8}$$

where ρ^{T_A} is the partial transpose on the subsystem A and $\|\cdot\|_1$ denotes the trace (L_1) norm. The logarithmic negativity coincides with the Rényi-1/2 entanglement entropy for the pure state and is nonincreasing with the error rate of the channel, a requirement that any measure of entanglement must satisfy [41,42]. The logarithmic negativity was previously used in the study of ground-state topological phases [43–45] and more recently for detecting topological order in finite temperature Gibbs states [27,28].

We expect that the universal topological contribution to the entanglement [46,47] will survive in the corrupted mixed state below a critical error rate and can be captured by the logarithmic negativity. Thus, the conjectured form of this quantity is

$$\mathcal{E}_A = c|\partial A| - \gamma_N + \ldots, \tag{9}$$

where $|\partial A|$ is the circumference of the region A, c is a nonuniversal coefficient, and ellipsis denotes terms that vanish in the limit $|\partial A| \to \infty$. The constant term γ_N is the *topological entanglement negativity* of a simply connected subregion. It is argued to be a topological invariant that cannot come from local contributions to the entanglement due to the conversion property $\mathcal{E}_A = \mathcal{E}_{\bar{A}}$, i.e., negativity of a subsystem is equal to that of the complement [27,48]. Let us repeat the argument here for the reader's convenience. We assume that the nontopological part, arising from local contributions, can be written as an integral along the entanglement cut, $\mathcal{E}_{A,\mathrm{local}} = \int_{\partial A} f(\kappa, \partial\kappa) dl$, where $f(\kappa, \partial\kappa)$ depends on the extrinsic curvature κ of the cut. For a smooth and large entanglement cut, one can perform a Taylor expansion $f(\kappa, \partial\kappa) = f_0 + f_1\kappa + \ldots$, which integrates to $c|\partial A| + c_1 + c_2|\partial A|^{-1} + \ldots$. Notably, the extrinsic curvature changes its sign when transforming A to $\bar{A}$, necessitating the vanishing of all odd-order terms to ensure $\mathcal{E}_A = \mathcal{E}_{\bar{A}}$. In particular, $f_1 = 0$ and $c_1 = 0$. Thus, local contributions cannot produce a constant term in the negativity. In contrast, the von Neumann entropy of a subregion in the error-corrupted mixed state exhibits a volume-law scaling, and its constant piece is not topological because $S_A \neq S_{\bar{A}}$.

To facilitate the calculation of the negativity, we consider the Rényi negativity of even order

$$\mathcal{E}_A^{(2n)}(\rho) := \frac{1}{2-2n}\log\frac{\mathrm{tr}(\rho^{T_A})^{2n}}{\mathrm{tr}\rho^{2n}}. \tag{10}$$

The logarithmic negativity is recovered in the limit $2n \to 1$. Here, we choose a particular definition of the Rényi negativity such that it exhibits an area-law scaling in the corrupted state. In Sec. III, we show explicitly that in the Toric code the topological part $\gamma_N^{(2n)}$ of the Rényi negativity takes a quantized value $\log 2$ in the phase where the quantum memory is retained and vanishes otherwise.

To summarize, we expect the topological negativity takes the same universal value as the topological entanglement entropy in the uncorrupted ground state and drops sharply to a lower value at a critical error rate. It is *a priori* not clear, however, that the transition in the negativity must occur at the same threshold as that marks the transition of the other two diagnostics we discussed. In Sec. III we show, through mapping to a statistical mechanical model that, in the example of the Toric code, a single phase transition governs the behavior of all three diagnostics.

III. EXAMPLE: TORIC CODE UNDER BIT-FLIP AND PHASE ERRORS

In this section, we use the three information-theoretical diagnostics to probe the distinct error-induced phases in the 2D Toric code under bit-flip and phase errors. In particular, we develop 2D classical statistical mechanical models to analytically study the Rényi-n version of the diagnostics in this example. The statistical mechanical models involve $(n-1)$-flavor Ising spins and undergo ferromagnetic phase transitions as a function of error rates. We show that the three diagnostics map to distinct observables that all detect the ferromagnetic order and undergo the transition simultaneously. We remark that our results also apply to the planar surface code.

In Sec. III A, we introduce the Toric code and the error models. We derive the statistical mechanical models in Sec. III B and analyze the phase transition in Sec. III C.

TABLE I. Dictionary of the mapping. The Rényi-n version of the diagnostics of topological order in error-corrupted states and their corresponding observables in $(n-1)$-flavor Ising models are listed in the first and second columns, respectively. We consider 2D Toric code subject to one type of incoherent error (bit-flip or phase errors). The asymptotic behaviors of these diagnostics in the paramagnetic (PM) and ferromagnetic (FM) phases of the spin model are provided.

Diagnostics	Observable	PM	FM
$D^{(n)}$	Logarithm of order parameter correlation function	$O(\lvert i_l - i_r\rvert)$	$O(1)$
$I_c^{(n)}$	Related to the excess free energy for domain walls along noncontractible loops	$2\log 2$	0
$\mathcal{E}_A^{(2n)}$	Excess free energy for aligning spins on the boundary of A	$c\lvert\partial A\rvert/\xi - \log 2$	$c\lvert\partial A\rvert/\xi$

Section III D discusses the three diagnostics and their corresponding observables in the statistical mechanical models. See Table I for a summary. We discuss the replica limit $n \to 1$ in Sec. III E.

A. Toric code and error model

We consider the 2D Toric code on an $L \times L$ square lattice with periodic boundary conditions. This code involves $N = 2L^2$ physical qubits on the edges of the lattice, and its code space is given by the ground-state subspace of the Hamiltonian

$$H_{\mathrm{TC}} = -\sum_s A_s - \sum_p B_p, \tag{11}$$

where A_s and B_p are mutually commuting operators associated with vertices and plaquettes

$$A_s = \prod_{\ell\in \mathrm{star}(s)} X_\ell, \quad B_p = \prod_{\ell\in \mathrm{boundary}(p)} Z_\ell. \tag{12}$$

Here, X_ℓ and Z_ℓ denote the Pauli-X and Z operators on edge ℓ, respectively. The ground state satisfying $A_s\left|\Psi\right\rangle = B_p\left|\Psi\right\rangle = \left|\Psi\right\rangle$ is fourfold degenerate and can encode two logical qubits.

We consider specific error channels describing uncorrelated single-qubit bit-flip and phase errors

$$\begin{aligned}\mathcal{N}_{X,i}[\rho] &= (1-p_x)\rho + p_x X_i \rho X_i,\\ \mathcal{N}_{Z,i}[\rho] &= (1-p_z)\rho + p_z Z_i \rho Z_i,\end{aligned} \tag{13}$$

where the Pauli-X (Z) operator acting on the Toric code ground state creates a pair of m (e) anyons on the adjacent plaquettes (vertices), p_x and p_z are the corresponding error rates. The corrupted state reads

$$\rho = \mathcal{N}_X \circ \mathcal{N}_Z[\rho_0],$$

where $\mathcal{N}_{X(Z)} = \prod_i \mathcal{N}_{X(Z),i}$. We assume that the error rate is uniform throughout our discussion. We remark that the error channels in Eq. (13) do not create coherent superposition between states with different anyon configurations and are referred to as incoherent errors. Pauli-Y errors create anyons incoherently and can also be analyzed using our framework.

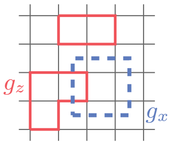
B. Statistical mechanical models

Here, we map the nth moment of the corrupted density matrix $\mathrm{tr}\rho^n$ to the partition function of the $(n-1)$-flavor Ising model. In this statistical mechanical model, one can analyze the singularity in the Rényi version of the three diagnostics, which will be presented in Sec. III D.

To begin, we consider the maximally mixed state in the ground-state subspace

$$\rho_0 = \frac{1}{4}\prod_s \frac{1+A_s}{2}\prod_p \frac{1+B_p}{2}. \tag{14}$$

We note that the choice of the ground state ρ_0 determines the boundary conditions in the resulting model and does not affect the location of the critical point. For our purpose here, it is convenient to write ρ_0 in a loop picture



$$\rho_0 = \frac{1}{2^N}\sum_{g_z} g_z \sum_{g_x} g_x\,, \tag{15}$$

where g_z and g_x are Z and X loops on the original and dual lattice given by the product of A_s and B_p operators, respectively. The summation runs over all possible loop configurations. In what follows, we will use $g_{x(z)}$ to denote both the operators and the loop configurations. The meaning will be clear in the context.

Two error channels act on the loop operators g_x, g_z by only assigning a real positive weight:

$$\mathcal{N}_{X,i}[g_z] = \begin{cases}(1-2p_x)g_z & Z_i \in g_z\\ g_z & Z_i \notin g_z\end{cases},$$

$$\mathcal{N}_{Z,i}[g_x] = \begin{cases}(1-2p_z)g_x & X_i \in g_x\\ g_x & X_i \notin g_x\end{cases}.$$

Thus, the corrupted state remains a superposition of loop operators

$$\rho = \frac{1}{2^N}\sum_{g_x,g_z} e^{-\mu_x|g_x|-\mu_z|g_z|} g_x g_z, \tag{16}$$

where $|g_{x(z)}|$ denotes the length of the loop, and $\mu_{x(z)} = -\log(1-2p_{z(x)})$ can be understood as the line tension.

Using Eq. (16), it is straightforward to see that the expectation values of operators, such as the Wilson loop and open string, behave smoothly as the error rate increases, in consistence with the general argument in Sec. II A.

Using this loop picture Eq. (16), we can write the nth moment as

$$\mathrm{tr}\rho^n = \frac{1}{2^{nN}} \sum_{\{g_x^{(s)},g_z^{(s)}\}} \mathrm{tr}\left(\prod_{s=1}^{n} g_x^{(s)} g_z^{(s)}\right) e^{\sum_s -\mu_x |g_x^{(s)}| - \mu_z |g_z^{(s)}|}, \tag{17}$$

where $g_{x(z)}^{(s)}$, $s = 1, 2, \ldots, n$ is the $X(Z)$ loop operator from the sth copy of density matrix. The product of loop operators in Eq. (17) has a nonvanishing trace only if the products of X and Z loops are proportional to identity individually, which leads to two independent constraints

$$g_a^{(n)} = \prod_{s=1}^{n-1} g_a^{(s)}, \quad a = x, z. \tag{18}$$

The nth moment factorizes into a product of two partition functions

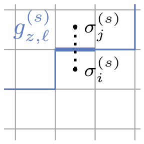
$$\mathrm{tr}\rho^n = \frac{1}{2^{(n-1)N}} \mathcal{Z}_{n,x} \mathcal{Z}_{n,z}, \tag{19}$$

where $\mathcal{Z}_{n,a} = \sum_{\{g_a^{(s)}\}} e^{-H_{n,a}}$ with $a = x, z$ is a statistical mechanical model that describes fluctuating $X(Z)$ loops with a line tension. The Hamiltonian takes the form

$$H_{n,a} = \mu_a \left(\sum_{s=1}^{n-1} |g_a^{(s)}| + \left| \prod_{s=1}^{n-1} g_a^{(s)} \right| \right). \tag{20}$$

Here, we have imposed the constraints (18), and the summation in each partition function runs over the loop configurations only in the first $n-1$ copies.

The loop model can be mapped to a statistical mechanical model of $n-1$ flavors of Ising spins with nearest-neighbor ferromagnetic interactions. The mapping is established by identifying the loop configuration $g_a^{(s)}$ with $s = 1, 2, \ldots, n-1$ with domain walls of Ising spins. Specifically, for a Z loop configuration on the original lattice, we associate a Ising spin configuration σ_i on the dual lattice such that



$$\left|g_{z,\ell}^{(s)}\right| = \left(1 - \sigma_i^{(s)} \sigma_j^{(s)}\right)/2,$$

where i,j are connected by the link dual to ℓ, and $|g_{z,\ell}^{(s)}|$ is a binary function that counts the support of loop on link ℓ. The total length of the loop is given by $|g_z^{(s)}| = \sum_\ell |g_{z,\ell}^{(s)}|$. Similarly, we can define the Ising spins on the original lattice that describe the X loop configuration on the dual lattice.

In terms of the Ising spins, the effective Hamiltonian is given by

$$H_{n,a} = -J_a \sum_{\langle i,j \rangle} \left(\sum_{s=1}^{n-1} \sigma_i^{(s)} \sigma_j^{(s)} + \prod_{s=1}^{n-1} \sigma_i^{(s)} \sigma_j^{(s)} \right) \tag{21}$$

with a ferromagnetic coupling $J_{x(z)} = -\log\sqrt{1 - 2p_{z(x)}}$. In what follows, we refer to this model as the $(n-1)$-*flavor Ising model*. We remark that the model exhibits a global symmetry $G^{(n)} = (\mathbb{Z}_2^{\otimes n} \rtimes \mathcal{S}_n)/\mathbb{Z}_2$, where $\mathcal{S}_n$ is the permutation symmetry over n elements. As is shown below, increasing the error rate the model undergoes a paramagnetic-to-ferromagnetic transition that completely breaks the $G^{(n)}$ symmetry.

C. Phase transitions

Here, we study the ferromagnetic transition in the $(n-1)$-flavor Ising model. The transition points depend on n and are determined using both analytical methods (e.g., Kramers-Wannier duality for $n = 2, 3$) and Monte-Carlo simulation (for $n = 4, 5, 6$, etc.). The results are presented in Fig. 2.

For $n = 2$, the statistical mechanical model is the standard square-lattice Ising model:

$$H_{2,a} = -2J_a \sum_{\langle i,j \rangle} \sigma_i \sigma_j. \tag{22}$$

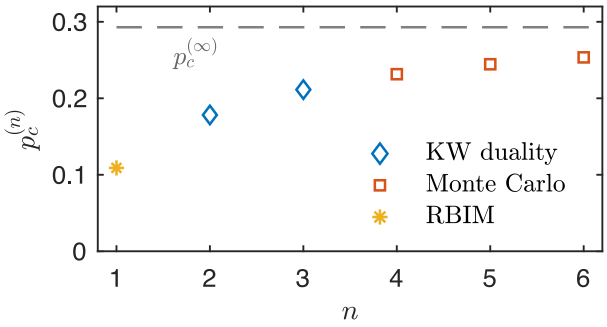


FIG. 2. Critical error rates for various Rényi index n. $p_c^{(2)} \approx 0.178$ and $p_c^{(3)} \approx 0.211$ are determined by the exact solution (blue diamonds). For $n \geqslant 4$, $p_c^{(n)}$ is determined by calculating the crossing of the Binder ratio for various system sizes via Monte Carlo (red squares). $p_c^{(n)}$ in the replica limit $n \to 1$ (the yellow star) is given by the critical point of random-bond Ising model (RBIM) in 2D, $p_c^{(1)} \approx 0.109$, as explained in Sec. III E. In the limit $n \to \infty$, the spin model is asymptotically decoupled Ising models with $p_c^{(\infty)} \approx 0.293$ (the gray dashed line).

The critical point is determined analytically by the Kramers-Wannier duality [49,50]

$$p_c^{(2)} = \frac{1}{2}\left(1 - \sqrt{\sqrt{2} - 1}\right) \approx 0.178. \tag{23}$$

For $n = 3$, the model becomes the Ashkin-Teller model on 2D square lattice along the $\mathcal{S}_4$ symmetric line. The Hamiltonian is

$$H_{3,a} = -J_a \sum_{\langle i,j\rangle} \sigma_i^{(1)}\sigma_j^{(1)} + \sigma_i^{(2)}\sigma_j^{(2)} + \sigma_i^{(1)}\sigma_i^{(2)}\sigma_j^{(1)}\sigma_j^{(2)}. \tag{24}$$

The model is equivalent to the standard four-state Potts model [51] with a critical point determined by the Kramers-Wannier duality

$$p_c^{(3)} = \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) \approx 0.211. \tag{25}$$

For $n \geqslant 4$, we are not aware of any exact solution and resort to the Monte Carlo simulation. To locate the transition point p_c, we consider the average magnetization per spin,

$$m := \frac{1}{(n-1)L^2} \sum_{s=1}^{n-1} \sum_i \sigma_i^{(s)}. \tag{26}$$

We calculate the magnetization square $\langle m^2\rangle$ and the Binder ratio $B = \langle m^4\rangle / \langle m^2\rangle^2$ numerically and display the results in Fig. 3. Assuming a continuous transition, we determine $p_c^{(n)}$ by the crossing point of $B(p, L)$ for various system sizes L and extract the critical exponents using the scaling ansatz $B(p, L) = \mathcal{F}_b((p - p_c)L^{1/\nu})$ and $\langle m^2\rangle(p, L) = L^{-2\beta/\nu}\mathcal{F}_m((p - p_c)L^{1/\nu})$. The analysis yields $p_c^{(4)} = 0.231$ for $n = 4$. However, the sharp drop of magnetization and the nonmonotonic behavior of $B(p, L)$ near $p_c^{(4)}$ hint at a possible first-order transition [52,53].

The critical error threshold p_c increases monotonically with n and is exactly solvable in the limit $n \to \infty$. In this case, the interaction among different flavors is negligible compared to the two-body Ising couplings. Thus, the critical point is asymptotically the same as that in the Ising model with coupling J_a and is given by

$$p_{a,c}^{(\infty)} = \frac{1}{2}\left(2 - \sqrt{2}\right) \approx 0.293. \tag{27}$$

D. Three diagnostics

The Rényi version of the three information-theoretic diagnostics, quantum relative entropy, coherent information, and topological entanglement negativity, translate into distinct physical quantities in the statistical mechanical model. We write these quantities explicitly below and show that all three detect the establishment of ferromagnetic order. Therefore, the transition in all three quantities is governed by the same critical point, a fact that is not evident before mapping to statistical mechanical models.

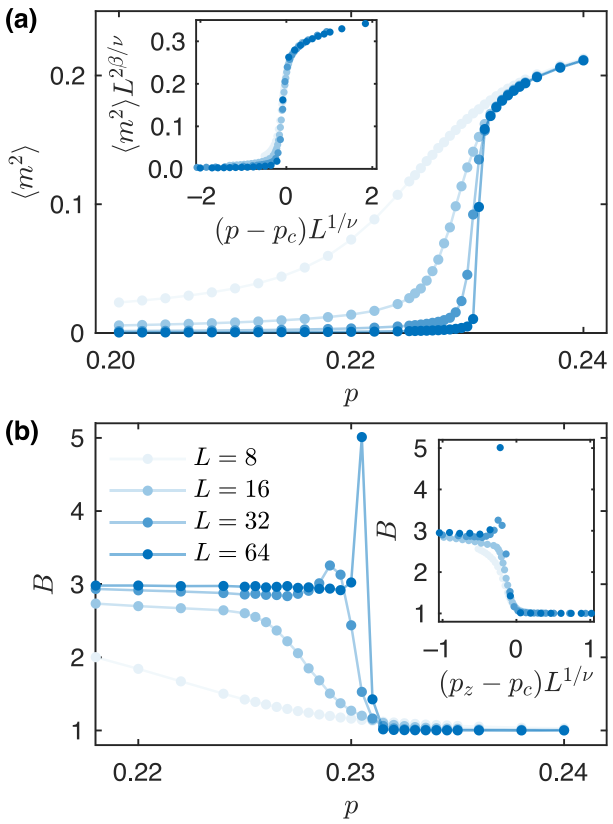


FIG. 3. Phase transition in the statistical mechanical model for $n = 4$. Magnetization (a) and Binder ratio (b) as a function of error rate p for various system sizes up to $L_x = L_y = L = 64$. The crossing of $B(p, L)$ yields $p_c = 0.231$. The exponents $\nu = 0.74$ and $\beta = 0.04$ are extracted from the finite-size scaling collapse in the insets. The results are averaged over 10^5 independent Monte Carlo measurements for each of 48 initial spin configurations.

1. Quantum relative entropy

We start with the Rényi version of the quantum relative entropy given by Eq. (4). Let ρ be the corrupted ground state of the Toric code, and $\rho_m = \mathcal{N}[|\Psi_m\rangle\langle\Psi_m|]$ where $|\Psi_m\rangle := w_m(\mathcal{C})|\Psi_0\rangle$ has a pair of m particles at the end of path $\mathcal{C}$. The phase errors do not change the distinguishability between the two states and can be safely ignored here. Only the statistical mechanical model for the Z loops (or Z spins) is relevant. Let i_ℓ and i_r denote the positions of two m particles, we show in Appendix A 1 that the Rényi relative entropy is mapped to a two-point function of the Ising spins

$$D^{(n)}(\rho||\rho_\alpha) = \frac{1}{1-n}\log\langle\sigma_{i_\ell}^{(1)}\sigma_{i_r}^{(1)}\rangle, \tag{28}$$

where $\sigma_j^{(1)}$ is the first flavor of the Ising spin at site j, and the subscription z is suppressed.

When the error rate is small and the system is in the paramagnetic phase, the correlation function decays exponentially, and thus $D^{(n)} = O(|i_\ell - i_r|)$, which grows linearly with the distance between i_ℓ and i_r. This indicates that the error-corrupted ground state and excited state remain distinguishable. When the error rate exceeds the critical value and the system enters the ferromagnetic phase, $D^{(n)}$ is of $O(1)$ due to the long-range order, which implies that the error-corrupted ground state and excited state are no longer distinguishable.

2. Coherent information

Next consider the Rényi version of the coherent information $I_c^{(n)}$ in Eq. (7). We let the two logical qubits in the system Q be maximally entangled with two reference qubits R. As detailed in Appendix A 2, $I_c^{(n)}$ can be mapped to the free energy cost of inserting domain walls along noncontractible loops that are related to the logical operators. More explicitly, let $\mathbf{d}_{al}$ with $a = x, z$ and $l = l_1, l_2$ be a $(n-1)$-component binary vector. Each component of $\mathbf{d}_{al}$ dictates the insertion of domain walls for $a = x, z$ spins along the noncontractible loop l, respectively, in $n-1$ copies of the Ising spins. Here, along the domain walls, the couplings between nearest-neighbor spins are flipped in sign and turned antiferromagnetic. Then, we have

$$I_c^{(n)} = \frac{1}{n-1} \sum_{a=x,z} \log \left(\sum_{\mathbf{d}_{a1}\mathbf{d}_{a2}} e^{-\Delta F_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}} \right) - 2\log 2, \tag{29}$$

where $\Delta F_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}$ is the free energy cost associated with inserting domain walls labeled by binary vectors $\mathbf{d}_{al}$, the sum runs over all possible $\mathbf{d}_{al}$.

When the error rate is small and the system is in the paramagnetic phase, the domain wall along a noncontractible loop costs nothing, i.e., $\Delta F_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})} = 0$. It follows that the corrupted state retains the encoded information, i.e., $I_c^{(n)} = 2\log 2$. When the error rate exceeds the critical value and the system enters the ferromagnetic phase, inserting a domain wall will have a free energy cost that is proportional to its length. Namely, $\Delta F_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}$ is proportional to the linear system size unless no defect is inserted. One can deduce $I_c^{(n)} = 0$ when the spin model for either Z or X loop undergoes a transition to the ferromagnetic phase, namely, the corrupted state corresponds to a classical memory. When both spin models are in the ferromagnetic phase, we have $I_c^{(n)} = -2\log 2$, indicating that the system is a trivial memory.

3. Topological entanglement negativity

The Rényi negativities of even order are given in Eq. (10). Let us specialize here to the Toric code with only phase errors. As shown in Appendix A 3, the 2nth Rényi negativity of a region A is given by

$$\mathcal{E}_A^{(2n)} = \Delta F_A, \tag{30}$$

where ΔF_A is the excess free energy associated with aligning a single flavor of Ising spins on the boundary ∂A in the same direction (illustrated in Fig. 4).

The excess free energy ΔF_A, or more precisely, its subleading term can probe the ferromagnetic transition in the statistical-mechanical model. The excess free energy has two contributions. The energetic part is always proportional to $|\partial A|$. The entropic part is attributed to the loss of degrees of freedom due to the constraint. In the paramagnetic phase, the Ising spins fluctuate freely above the scale of the finite correlation length ξ. Hence, enforcing each constraint removes $O(|\partial A|/\xi)$ degrees of freedom proportional to the circumference of A, which yields the leading term (area law). Importantly, there is still one residual degree of freedom, namely, the aligned boundary spins can fluctuate together, which results in a subleading term $\log 2$. Altogether, we have $\mathcal{E}_A^{(2n)} = c|\partial A|/\xi - \log 2$. Here, it is an interesting question to verify whether the prefactor c is universal or not [54], and we leave it for future study [55]. In the ferromagnetic phase, the finite correlation length ξ sets the scale of the critical region, below which the spins can fluctuate. Thus, imposing each constraint removes $O(|\partial A|/\xi)$ degrees of freedom. However, the aligned boundary spins should also align with the global magnetization resulting in a vanishing subleading term in the excess free energy. Hence, the negativity $\mathcal{E}_A^{(2n)}$ exhibits a pure area law without any subleading term.

To support our analytical argument, we also numerically calculate the Rényi-4 negativity (the Rényi-2 negativity is trivially zero) and show that the topological term $\gamma_N^{(4)}$ indeed exhibits distinct behaviors across the transition. We adopt the Kitaev-Preskill prescription to extract γ_N [46]. More specifically, we consider the subsystems A, B, C

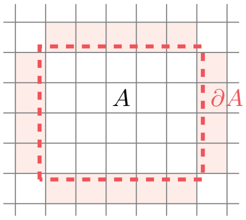
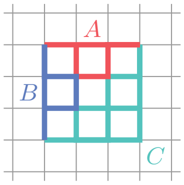


FIG. 4. Entanglement negativity between region A and its complement $\bar{A}$ corresponds to the excess free energy for aligning Ising spins on the boundary of A (pink plaquettes) pointing to the same direction.

depicted below, and γ_N is given by



$$-\gamma_N := \mathcal{E}_A + \mathcal{E}_B + \mathcal{E}_C + \mathcal{E}_{ABC} - \mathcal{E}_{AB} - \mathcal{E}_{BC} - \mathcal{E}_{AC}\,. \tag{31}$$

Our choice of the subsystems further simplifies the above expression to $-\gamma_N = 2\mathcal{E}_A - 2\mathcal{E}_{AC} + \mathcal{E}_{ABC}$ [56].

The result is presented in Fig. 5, where $\gamma_N^{(4)}$ approaches $\log 2$ and 0 for small and large p_z, respectively. The curves become steeper as the system size increases, which is consistent with the predicted step function in the thermodynamic limit. One can also observe a dip of $\gamma_N^{(4)}$ below zero. This phenomenon has also appeared in the numerical study of the topological entanglement entropy across transitions [13]. We believe that this dip is due to the finite-size effect, which might be more severe for information quantities with a large Rényi index n [57].

So far, we considered only a simply connected subregion. If A is not simply connected, that is, ∂A contains k disconnected curves (for example, the boundary of an annular region that contains two disconnected curves), then the constraints require only the Ising spins to align with other spins on the same boundary curve. In this case, the topological entanglement negativity is $k \log 2$. This is the same dependence on the number of disconnected components as in the topological entanglement entropy of ground states [47].

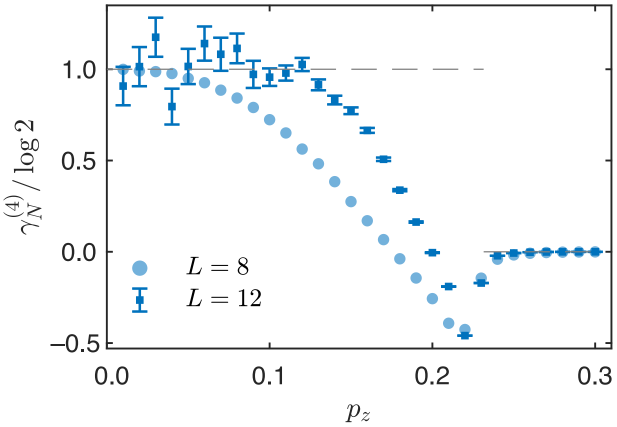
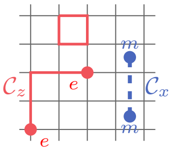


FIG. 5. Topological negativity $\gamma_N^{(4)}$ as a function of the phase error rate p_z. We consider the subsystems A, B, and C as in Eq. (31) and choose the side of the region ABC to be $L/4$. $\gamma_N^{(4)}$ approaches $\log 2$ and zero at small and large p_z, respectively. The curves become steeper as the system size L increases. The dashed line indicates the predicted behavior in the thermodynamic limit. The results are averaged over 10^7 independent Monte Carlo measurements from each of 48, 96 random initial spin configurations for $L = 8, 12$, respectively. The error bars for $L = 8$ are negligible and thus omitted.

E. $n \to 1$ limit, duality and connection to optimal decoding

In this subsection, we determine p_c in the limit $n \to 1$ via a duality between the statistical mechanical model established in Sec. III B and the 2D random bond Ising model (RBIM) along the Nishimori line [58]. The RBIM is also known to govern the error threshold of the optimal decoding algorithm for the 2D Toric code with incoherent errors [10]. The duality shows that the decoding threshold indeed saturates the upper bound given by the threshold in our information theoretical diagnostics. This duality was derived before via a binary Fourier transformation [59,60]. Here, it follows naturally from two distinct expansions of the error-corrupted state.

The statistical mechanical model in Sec. III B is based on the loop picture in Eq. (15). Here, we work in an alternative error-configuration picture, writing the error-corrupted state as



$$\rho = \sum_{\mathcal{C}_x, \mathcal{C}_z} P(\mathcal{C}_x) P(\mathcal{C}_z) Z^{\mathcal{C}_z} X^{\mathcal{C}_x} \rho_0 X^{\mathcal{C}_x} Z^{\mathcal{C}_z}\,, \tag{32}$$

where $\mathcal{C}_z$ ($\mathcal{C}_x$) denotes the error string on the original (dual) lattice. The error string creates error syndromes, i.e., e (m) anyons, on the boundary $\partial\mathcal{C}_z$ ($\partial\mathcal{C}_x$). The probability for each string configuration is

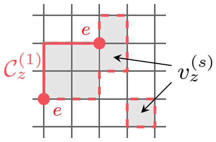
$$P(\mathcal{C}_a) = p_a^{|\mathcal{C}_a|}(1-p_a)^{N-|\mathcal{C}_a|}, \tag{33}$$

where $|\mathcal{C}_a|$ with $a = x, z$ denotes the total length of the error string, and N is the total number of qubits.

The expansion in error configurations allows writing the nth moment as

$$\mathrm{tr}\rho^n = \sum_{\{\mathcal{C}_x^{(s)}, \mathcal{C}_z^{(s)}\}} \prod_{s=1}^{n} P\left(\mathcal{C}_x^{(s)}\right) P\left(\mathcal{C}_z^{(s)}\right) \times \mathrm{tr}\left(\prod_{s=1}^{n} Z^{\mathcal{C}_z^{(s)}} X^{\mathcal{C}_x^{(s)}} \rho_0 X^{\mathcal{C}_x^{(s)}} Z^{\mathcal{C}_z^{(s)}}\right). \tag{34}$$

We again choose ρ_0 to be the maximally mixed state in the logical space. The trace is nonvanishing only when the error strings in two consecutive copies differ by a closed loop. Thus, the error strings in the $2, \ldots, n$th copies are related to that in the first copy via



$$\mathcal{C}_a^{(s+1)} = \mathcal{C}_a^{(1)} + \partial v_a^{(s)} + l_1^{d_{a1}^{(s)}} + l_2^{d_{a2}^{(s)}}, \quad s = 1, \ldots, n-1, \tag{35}$$

where $v_a^{(s)}$ is a set of plaquettes on the original (or dual) lattice, its boundary $\partial v_a^{(s)}$ consists only of homologically trivial loops. Two strings can also differ by a noncontractible loop l_1, l_2 on the torus indicated by the binary variables $d_{a1}^{(s)}, d_{a2}^{(s)} = 0, 1$. Noticing the decoupling between Z and X, we have

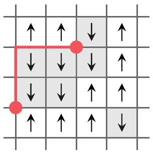
$$\begin{aligned} \mathrm{tr}\rho^n &= \mathcal{Z}'_{n,z}\mathcal{Z}'_{n,x}, \\ \mathcal{Z}'_{n,a} &= \sum_{\mathbf{d}_a} \sum_{\mathcal{C}_a^{(1)}} P\left(\mathcal{C}_a^{(1)}\right) \\ &\quad \times \sum_{\{v_a^{(s)}\}} \prod_{s=1}^{n-1} P\left(\mathcal{C}_a^{(1)} + \partial v_a^{(s)} + l_1^{d_{a1}^{(s)}} + l_2^{d_{a2}^{(s)}}\right), \end{aligned} \tag{36}$$

where we denote the collection of $(d_{a1}^{(s)}, d_{a2}^{(s)})$ for $s = 1, 2, \ldots, n-1$ as a binary vector $\mathbf{d}_a$. Comparing the above expression with Eq. (19), we have

$$\mathcal{Z}_{n,x} = 2^{\frac{(n-1)N}{2}} \mathcal{Z}'_{n,z}, \quad \mathcal{Z}_{n,z} = 2^{\frac{(n-1)N}{2}} \mathcal{Z}'_{n,x}. \tag{37}$$

In the following, we focus on $\mathcal{Z}'_{n,z}$ and suppress the subscripts for clarity. The analysis of $\mathcal{Z}'_{n,x}$ is similar.

We now interpret $\mathcal{Z}'_n$ as a partition function of Ising spins that is related to the replicated RBIM. We first introduce $n-1$ flavors of Ising spins on the plaquettes to represent $v^{(s)}, s = 1, \ldots, n-1$. The Ising domain wall represents $\partial v^{(s)}$ as shown below.



Next, we identify the probability of error strings with the Boltzmann weight of Ising spin configurations. Effectively, the spins of the same flavor have nearest-neighbor antiferromagnetic interactions if their link crosses the path $\mathcal{C}^{(1)}$ or $l_{1(2)}$ when $d_{1(2)}^{(s)} = 1$; the interaction is ferromagnetic otherwise. Specifically,

$$\mathcal{Z}'_n = ((1-p)p)^{N/2} \sum_{\{\eta_{ij}\}} P(\{\eta_{ij}\}) \sum_{\mathbf{d}} \sum_{\tau^{(s)}} e^{-\mathsf{H}_n(\eta_{ij}, \mathbf{d})}, \tag{38}$$

where

$$\mathsf{H}_n(\eta_{ij}, \mathbf{d}) = -J \sum_{s=1}^{n-1} \sum_{\langle ij \rangle} \xi_{ij}^{(s)}(\mathbf{d}) \eta_{ij} \tau_i^{(s)} \tau_j^{(s)}. \tag{39}$$

Here, J depends on p and satisfies the Nishimori condition $e^{-2J} = p/(1-p)$ [10,58]. Both η_{ij} and $\xi_{ij}^{(s)}(\mathbf{d})$ take the value ±1. The random variable η_{ij} takes the -1 value along $\mathcal{C}^{(1)}$, which can be interpreted as a random sign of bond coupling. The variable $\xi_{ij}^{(s)}(\mathbf{d}) = -1$ along noncontractible loops $l_{1(2)}$ when $d_{1(2)}^{(s)} = 1$, which can be interpreted as a defect in the spin model. The above expression allows writing $\mathcal{Z}'_n = \overline{\mathcal{Z}_{\mathrm{RBIM}}^{n-1}}$ as the disorder-averaged partition function of $n-1$ copies of RBIM along the Nishimori line.

The replicated RBIM in the error configuration picture and the spin model in the loop picture are derived from the n-th moment of the *same* error corrupted state. Therefore, they are dual to each other and share the same critical error rate for all replica indices. Note that the replicated RBIM exhibits two phases, a ferromagnetic and a paramagnetic phase at small and large error rates, respectively. The phase diagram is exactly opposite to that of the spin model in its dual picture, which is a common feature of Kramers-Wannier dualities [61].

In the replica limit $n \to 1$, the replicated RBIM reduces to the RBIM derived for the optimal quantum error-correction algorithm [10] and undergoes an ordering transition at $p_c = 0.109$ [62,63]. This implies that all three diagnostics should also undergo the transition at the same p_c in the replica limit and confirms that the optimal decoding threshold saturates the upper bound in Eq. (6).

IV. DISCUSSION

In this work, we introduced information-theoretic diagnostics of error-corrupted mixed states $\rho = \prod_i \mathcal{N}_i[\rho_0]$, which probe their intrinsic topological order and capacity for protecting quantum information. We focused on a concrete example, where ρ_0 is in the ground-state subspace of the Toric code and $\mathcal{N}_i$ describes the bit-flip and phase errors. We noted that the nth moment $\mathrm{tr}\rho^n$ can be written as the partition function of a 2D classical spin model, that is dual to the (replicated) random-bond Ising model along the Nishimori line, which is used to establish the following results. We consider three complementary diagnostics, quantum relative entropy, coherent information, and topological entanglement negativity, which are mapped to different observables in the spin model and shown to undergo a transition at the same critical error rate. Generally speaking, this critical error rate is an upper bound for the error threshold that can be achieved by any decoding algorithm. The aforementioned duality implies that the critical error rate identified here is exactly saturated by the error threshold of the optimal decoding algorithm for the Toric code proposed by Dennis *et al.* [10]. This result unveils a connection between the breakdown of topological quantum memory and a transition in the mixed-state topological order, and also provides physical interpretation for the decoding transition.

We have focused on Toric code with incoherent errors. It will be interesting to generalize the discussion to coherent errors that create anyons with coherence, e.g., amplitude damping or unitary rotations [64–67]. In these cases, one has to concatenate coherent errors and dephasing channels that mimic the syndrome measurement in order to make better contact to quantum error correction based on that syndrome measurement. It is also interesting to further consider non-Abelian quantum codes [68–70].

It might be surprising that the intrinsic properties of the 2D error-corrupted quantum states are captured by 2D *classical* statistical-mechanical models. In Appendix B, we give a brief discussion on $\mathbb{Z}_N$ Toric code with specific incoherent errors and show that this is also the case. A more general perspective is the so-called *errorfield double formalism*, which is proposed by the same authors. It follows from this general formalism that the intrinsic properties of the 2D error-corrupted states can always be captured by a 1+1D quantum model. Details will be reported elsewhere [71].

In the Toric code and other topological codes with local errors, the statistical mechanical model for the optimal decoding algorithm always satisfies the Nishimori condition [10,18–22,72]. One salient feature of the statistical mechanical model on the Nishimori line is an enlarged $\mathcal{S}_n$ symmetry in the replicated model of $n-1$ replicas [73,74]. In our analysis of intrinsic mixed-state transition, the $(n-1)$th replicated model actually corresponds to the nth moment $\mathrm{tr}\rho^n$, where the invariance under permuting n copies of the density matrix naturally gives rise to the $\mathcal{S}_n$ symmetry. This offers an alternative perspective on the occurrence of Nishimori physics in the context of optimal decoding.

As we have commented in Sec. II A, the error-induced transition acquires a different nature from the thermal transition in finite-temperature topological order. This distinction suggests a hierarchy of topological transitions in general mixed states. For example, it suffices to use physical observables (linear in the density matrix) to detect the thermal transition, while it requires at least second Rényi quantities (quadratic in the density matrix) to detect the error-induced transition. It is interesting to explore more exotic topological transitions in mixed states that are detectable only by nonlinear functions of the density matrix of even higher orders, such as the entanglement Hamiltonian.

The above task is intimately related to the goal of classifying mixed-state topological order. A suitable definition of mixed-state topological order should be both operationally meaningful and also identify computable topological invariants. Our discussion, which focuses on the error-corrupted mixed states, represents one particular aspect of this more general question. Here, the coherent information provides the operational definition, namely, a locally corrupted state is in a different phase if QEC is impossible, while the topological entanglement negativity is believed to provide a computable topological invariant that diagnoses the present transition. However, note that both the local error channel and QEC process are generally nonunitary, for which the Lieb-Robinson bound does not apply. Therefore, understanding the role of locality is key to obtaining a general notion of equivalence classes of mixed states. Similarly, a more general justification of topological negativity and its universality, in the sense of establishing its invariance under the application of local quantum channels at a certain place, is left for future work. The main difficulty comes from understanding how local perturbations affect the spectrum of a partially transposed density matrix, which is an interesting problem in its own right and is left to future work.

ACKNOWLEDGMENTS

We thank Meng Cheng, Soonwon Choi, Mikhail Lukin, Nishad Maskara, Karthik Siva, Tomohiro Soejima for helpful discussions, and Tarun Grover for useful comments on the paper. A.V. was funded by the Simons Collaboration on Ultra-Quantum Matter, which is a grant from the Simons Foundation (651440, A.V.). A.V. and R.F. further acknowledge support from NSF-DMR 2220703. Support is also acknowledged from the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator (EA). Y.B. was supported in part by NSF QLCI program through Grant No. OMA-2016245. This work is funded in part by a QuantEmX Grant from ICAM and the Gordon and Betty Moore Foundation through Grant No. GBMF9616 to Ruihua Fan and Yimu Bao. This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley.

Note added.—Recently, we became aware of an independent work [75], which is broadly related and will appear on arXiv on the same day. We thank the authors for informing us of their work in advance.

APPENDIX A: DETAILS OF THE MAPPING

In this section, we detail the mapping between the three diagnostics and observables in the statistical-mechanical models.

1. Quantum relative entropy

We here explicitly show that the Rényi quantum relative entropy is related to the correlation function in the classical spin model. Specifically, we consider the relative entropy between the error-corrupted ground state and an excited state $|\Psi_m\rangle := w_m(\mathcal{C})\,|\Psi_0\rangle$ with a pair of m particles created at the end of path $\mathcal{C}$.

First, we write down the error-corrupted state ρ_m in the loop representation

$$\rho_m = \frac{1}{2^N}\sum_g \mathrm{sgn}\left(g_z, X^{\mathcal{C}}\right) g_z g_x e^{-\mu_x|g_x|-\mu_z|g_z|}, \tag{A1}$$

where the commutation relation between the loop operator and the string operator is accounted by $\mathrm{sgn}(g_z, X^{\mathcal{C}})$; the sign function equals $+1$ when g_z and $X^{\mathcal{C}}$ commute and -1 otherwise. The above expression allows one to write $\mathrm{tr}\rho\rho_m^{n-1}$ as

$$\mathrm{tr}\rho\rho_m^{n-1} = \frac{\mathcal{Z}_{n,x}}{2^{(n-1)N}}\sum_{\{g_z^{(s)}\}} \mathcal{O}_D^{(n)} e^{-H_{n,z}}, \tag{A2}$$

where $\mathcal{O}_D^{(n)}$ denotes the product of sign functions in $n-1$ copies of ρ_m

$$\mathcal{O}_D^{(n)} = \mathrm{sgn}\left(g_z^{(1)}, X^{\mathcal{C}}\right). \tag{A3}$$

Here, we have used the constraint $g_z^{(1)} = \prod_{s=2}^n g_z^{(s)}$ for nonvanishing trace in the loop representation. Using this expression, the nth Rényi relative entropy takes the form

$$D^{(n)}(\rho||\rho_m) = \frac{1}{1-n}\log\langle\mathcal{O}_D^{(n)}\rangle. \tag{A4}$$

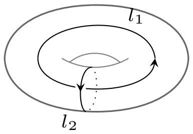
Our next step is to express the observable $\langle\mathcal{O}_D^{(n)}\rangle$ in terms of the Ising spins. In the spin model, the closed loop $g_z^{(1)}$ is identified with the domain wall of $\sigma_i^{(1)}$, and the Ising spins on two sides of $g_z^{(1)}$ antialign. Thus, $\sigma_{i_l}^{(1)}$ and $\sigma_{i_r}^{(1)}$ on the two ends of the open string $\mathcal{C}$ is aligned if $g_z^{(1)}$ crosses $\mathcal{C}$ for even number of times and is antialigned otherwise. The parity of the crossing is exactly measured by the sign function $\mathrm{sgn}(g_z^{(1)}, X^{\mathcal{C}})$. Hence, the observable $\langle\mathcal{O}_D^{(n)}\rangle$ maps to the correlation function

$$\langle\mathcal{O}_D^{(n)}\rangle = \langle\sigma_{i_l}^{(1)}\sigma_{i_r}^{(1)}\rangle. \tag{A5}$$

2. Coherent information

We now develop a spin-model description for the Rényi coherent information $I_c^{(n)}$ in Eq. (7). In the definition of coherent information, the system density matrix ρ_Q is the error-corrupted state ρ in Sec. III B, and its nth moment is mapped to the partition function of the $(n-1)$-flavor Ising model on the torus. Here, we show that the nth moment of ρ_{RQ} maps to the partition function of the same model with defects (domain walls) inserted along large loops on the torus.

First, we write down the initial state of the system Q and the reference R. We consider two reference qubits and two logical qubits in the ground-state subspace, and maximally entangle them in a Bell state. Let $s_l^{a=x,z}$ be the Pauli operator of two reference qubits, and $\bar{g}_{al}$ be the four logical operators



$$\bar{g}_{zl} := \prod_{\ell\in l} Z_\ell\,, \qquad \bar{g}_{xl} := \prod_{\ell\in l^*} X_\ell\,, \tag{A6}$$

where $l = l_{1,2}$ and $l^* = l^*_{1,2}$ are on the original and dual lattice. We consider the Bell state prepared as the $+1$ eigenstate of stabilizers $\bar{g}_{zl}s_l^z$ and $\bar{g}_{xl}s_l^x$, and write the initial density matrix for the system and reference as

$$\rho_{0,RQ} = \prod_{l=l_1,l_2}\prod_{a=x,z}\frac{1+\bar{g}_{al}s_l^a}{2}\prod_s\frac{1+A_s}{2}\prod_p\frac{1+B_p}{2}. \tag{A7}$$

Here, we again work in the loop picture of $\rho_{0,RQ}$, and further factorize the density matrix into a product

$$\rho_{0,RQ} = \frac{1}{2^{N+2}}\Gamma^x_{0,RQ}\Gamma^z_{0,RQ}, \tag{A8}$$

where $\Gamma^a_{0,RQ}$ is a summation of $a = x, z$ loops and takes the form

$$\Gamma^a_{0,RQ} = \sum_{g_a} g_a \prod_{l=l_1,l_2}\left(1+\bar{g}_{a,l}s_l^a\right). \tag{A9}$$

In the error-corrupted state ρ_{RQ}, the X and Z error channels act on $\Gamma^z_{0,RQ}$ and $\Gamma^x_{0,RQ}$, respectively, giving rise to $\rho_{RQ} = \Gamma^x_{RQ}\Gamma^z_{RQ}/2^{N+2}$ with

$$\Gamma^a_{RQ} = \sum_{g_a}\sum_{d_{al}=0,1} e^{-\mu_a\left|\prod_{l=l_1,l_2}(\bar{g}_{al})^{d_{al}}g_a\right|} g_a \prod_{l=l_1,l_2}(\bar{g}_{al}s_l^a)^{d_{al}}, \tag{A10}$$

where d_{al} is a binary variable indicating whether the loop operator in the summation acts on the noncontractible loop l of the torus.

Our next step is to write down the nth moment of ρ_{RQ} in the loop picture

$$\mathrm{tr}\rho_{RQ}^n = \frac{1}{2^{n(N+2)}}\mathrm{tr}\left(\left(\Gamma_{RQ}^x\right)^n\left(\Gamma_{RQ}^z\right)^n\right), \tag{A11}$$

where each $\Gamma_{RQ}^{x(z)}$ is a sum over all possible $X\,(Z)$ loop operators with positive weights. The product of loop operators from n copies has a nonvanishing trace only if the product is identity. This imposes the constraint on loop configurations and allows expressing the nth moment as a sum of partition functions

$$\mathrm{tr}\rho_{RQ}^n = \frac{1}{2^{(n-1)(N+2)}}\prod_{a=x,z}\sum_{\mathbf{d}_{a1}\mathbf{d}_{a2}}\mathcal{Z}_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}, \tag{A12}$$

where $\mathbf{d}_{al}$ with $l=1,2$ is a $(n-1)$-component binary vector, the sum runs over all possible $\mathbf{d}_{al}$, and $\mathcal{Z}_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})} = \sum_{\{g_a^{(s)}\}} e^{-H_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}}$ is the partition function with an effective Hamiltonian

$$\begin{aligned}H_{n,a}^{(\mathbf{d}_{1a},\mathbf{d}_{2a})} = {}& \mu_a\sum_{s=1}^{n-1}\left|(\bar{g}_{a1}^{(s)})^{d_{a1,s}}(\bar{g}_{a2}^{(s)})^{d_{a2,s}}g_a^{(s)}\right| \\ & + \mu_a\left|\prod_{s=1}^{n-1}(\bar{g}_{a1}^{(s)})^{d_{a1,s}}(\bar{g}_{a2}^{(s)})^{d_{a2,s}}g_a^{(s)}\right|.\end{aligned} \tag{A13}$$

Here, $d_{al,s}$ denotes the sth component of vector $\mathbf{d}_{al}$.

The loop model in Eq. (A13) can be identified with a classical spin model similar to Eq. (21). However, there is an important difference due to the presence of the homologically nontrivial loop $\bar{g}_{al}^{(s)}$. Here, we interpret the homologically trivial loop $g_a^{(s)}$ as the Ising domain wall and $\bar{g}_{al}^{(s)}$ as a defect along the noncontractible loop. The defect corresponds to flipping the sign of Ising coupling along a large loop. Specifically, for $Z\,(X)$ loops on the original lattice, we introduce Ising spin on the plaquettes (vertices) such that

$$\left|(\bar{g}_{a1})_\ell^{d_{a1,s}}(\bar{g}_{a2})_\ell^{d_{a2,s}}g_{a,\ell}^{(s)}\right| = \frac{1-(-1)^{\lambda_\ell^{(s)}}\sigma_i^{(s)}\sigma_j^{(s)}}{2}, \tag{A14}$$

where i,j are connected by the link ℓ, and $\lambda_\ell^{(s)} = |(\bar{g}_{a1})_\ell^{d_{a1,s}}(\bar{g}_{a2})_\ell^{d_{a2,s}}|$ is binary variable that denotes whether the defect goes through the link ℓ. This results in an effective Hamiltonian

$$\begin{aligned}H_{n,a}^{(\mathbf{d}_{1a},\mathbf{d}_{2a})} = {}& -J_a\sum_{\langle i,j\rangle}\sum_{s=1}^{n-1}(-1)^{\lambda_\ell^{(s)}}\sigma_i^{(s)}\sigma_j^{(s)} \\ & + \prod_{s=1}^{n-1}(-1)^{\lambda_\ell^{(s)}}\sigma_i^{(s)}\sigma_j^{(s)}.\end{aligned} \tag{A15}$$

Hence, $\mathcal{Z}_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}$ becomes the partition function of the classical spin model with defects inserting along the noncontractible loops labeled by binary vectors $\mathbf{d}_{al}$.

The mapping developed above allows a spin-model description for the nth Rényi coherent information $I_c^{(n)}$. The nth moment of ρ_Q is identified with the partition function with no defect, i.e., $\mathrm{tr}\rho_Q^n = \mathcal{Z}_{n,x}^{(\mathbf{0},\mathbf{0})}\mathcal{Z}_{n,z}^{(\mathbf{0},\mathbf{0})}/2^{(n-1)N}$. Therefore, we have

$$I_c^{(n)} = \frac{1}{n-1}\sum_{a=x,z}\log\frac{\sum_{\mathbf{d}_{a1}\mathbf{d}_{a2}}\mathcal{Z}_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}}{2^{n-1}\mathcal{Z}_{n,a}^{(\mathbf{0},\mathbf{0})}}. \tag{A16}$$

Thus, the Rényi coherent information is associated with the excess free energy of inserting defects along noncontractible loops

$$\Delta F_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})} := -\log\left(\mathcal{Z}_{n,a}^{(\mathbf{d}_{a1},\mathbf{d}_{a2})}/\mathcal{Z}_{n,a}^{(\mathbf{0},\mathbf{0})}\right). \tag{A17}$$

3. Entanglement negativity

Here, we show that the Rényi negativity in the error-corrupted state maps to the excess free energy for aligning spins in the statistical mechanical model. Specifically, we consider the case when only one type of error, e.g., bit-flip errors, is present.

The first step is to write down the partially transposed density matrix ρ^{T_A}. We again work in the loop representation, where the error-corrupted state is expressed as a sum of Pauli strings $g = g_x g_z$ in Eq. (16). The Pauli string g is invariant under the partial transpose up to a sign factor $y_A(g) = (-1)^{N_Y}$ depending on the number N_Y of Pauli-Y operators inside the subsystem A. Hence,

$$\rho^{T_A} = \frac{1}{2^N}\sum_g y_A(g)e^{-\mu_x|g_x|-\mu_z|g_z|}g. \tag{A18}$$

Using the above expression, one can write down the nth moment of ρ^{T_A}

$$\mathrm{tr}\left(\rho^{T_A}\right)^n = \frac{1}{2^{(n-1)N}}\sum_{\{g^{(s)}\}}\mathcal{O}_N^{(n)}e^{-H_{n,x}-H_{n,z}}. \tag{A19}$$

Here, similar to $\mathrm{tr}\rho^n$, the trace imposes a constraint on the loop operators $g^{(s)}$, and the summation runs over $g^{(s)}$ only in the first $n-1$ copies. The sign factors collected from the partial transpose in each copy are combined in $\mathcal{O}_N^{(n)}$,

$$\mathcal{O}_N^{(n)} = \left[\prod_{s=1}^{n-1}y_A\left(g^{(s)}\right)\right]y_A\left(\prod_{s=1}^{n-1}g^{(s)}\right). \tag{A20}$$

Equation (A19) allows expressing the $2n$th Rényi negativity in terms of the expectation value of $\mathcal{O}_{2n}$:

$$\mathcal{E}_A^{(2n)} = \frac{1}{2-2n}\log\left\langle\mathcal{O}_N^{(2n)}\right\rangle. \tag{A21}$$

Yet, analyzing the number of Pauli-Y operators in Eq. (A20) is a formidable task. Moreover, the observable $\mathcal{O}_N^{(n)}$ derived from the partial transpose should be a basis-independent quantity. Indeed, one can express $O_N^{(n)}$ in terms of loop configurations

$$\mathcal{O}_N^{(n)} = \prod_{r=1}^{n-2} \mathrm{sgn}_A\left(\prod_{s=1}^{r} g^{(s)}, g^{(r+1)}\right)$$
$$= \prod_{r=2}^{n-1}\prod_{s=1}^{r-1} \mathrm{sgn}_A\left(g^{(s)}, g^{(r)}\right). \qquad \text{(A22)}$$

Here, we use the property

$$y_A(g)y_A(h) = y_A(gh)\,\mathrm{sgn}_A(g,h), \qquad \text{(A23)}$$

where the sign function $\mathrm{sgn}_A(g,h) = \pm 1$ depending on the commutation relation between the support of Pauli string g and h on subsystem A:

$$\mathrm{sgn}_A(g,h) = \begin{cases} 1 & [g_A,\ h_A] = 0 \\ -1 & \{g_A,\ h_A\} = 0 \end{cases}. \qquad \text{(A24)}$$

In the second equality of Eq. (A22), we use the property of sign function

$$\mathrm{sgn}_A(g_1 g_2, g_3) = \mathrm{sgn}_A(g_1, g_3)\,\mathrm{sgn}_A(g_2, g_3). \qquad \text{(A25)}$$

In the Toric code, the operator g further factorizes into $g = g_x g_z$, where g_x, g_z are closed-loop operators of Pauli X and Z, respectively. The sign function between two such loop operators g and h reduces to

$$\mathrm{sgn}_A(g,h) = \mathrm{sgn}_A(g_x, h_z)\,\mathrm{sgn}_A(g_z, h_x). \qquad \text{(A26)}$$

We then arrive at

$$\mathcal{O}_N^{(n)} = \prod_{s,r=1,s\neq r}^{n-1} \mathrm{sgn}_A\left(g_x^{(s)}, g_z^{(r)}\right). \qquad \text{(A27)}$$

To develop an analytic understanding of the observable $\mathcal{O}_N^{(n)}$ and how it detects the ferromagnetic transition, we first consider the situation when only X or Z error is present. In this case, we show that $\log\langle\mathcal{O}_N^{(n)}\rangle$ exactly maps to the excess free energy of spin pinning and sharply distinguish the two phases. After that, we discuss the general situation when both types of error are present.

We here consider the case when only X errors are present, namely $p_z = 0$ and $\mu_x = 0$. The vanishing X-loop tension indicates that $H_{n,x}$ is in the paramagnetic phase, and the domain walls g_x of arbitrary sizes occur with the same probability. Thus, we can perform an exact summation over all possible g_x and obtain

$$\mathrm{tr}(\rho^{T_A})^n = \frac{1}{2^{(n-1)N}} \sum_{\{g_z^{(s)}\}} \mathcal{O}_{N,z}^{(n)} e^{-\mu_z H_{n,z}}, \qquad \text{(A28)}$$

where $\mathcal{O}_{N,z}^{(n)} = \sum_{\{g_x^{(s)}\}} \mathcal{O}_N^{(n)}$. The summation in $\mathcal{O}_{N,z}^{(n)}$ is nonvanishing only if the sign functions in Eq. (A27) for different $g_x^{(s)}$ interfere constructively. This yields a constraint on the $g_z^{(s)}$

$$\mathcal{O}_{N,z}^{(n)} = \prod_{r=1}^{n-1} N_{g_x} \delta_{h^{(r)}(A)}, \qquad \text{(A29)}$$

where $h^{(r)} = \prod_{s=1,s\neq r}^{n-1} g_z^{(s)}$, the Kronecker δ function $\delta_{h^{(r)}(A)}$ takes the value unity only if the support of $h^{(r)}$ on subsystem A is a closed loop and equals zero otherwise, and N_{g_x} is an unimportant prefactor that denotes the number of possible g_x in each copy. The $n-1$ δ function constraints are independent for odd n, whereas for even n they give rise to only $n-2$ independent constraints as $\prod_{r=1}^{n-1} h^{(r)} = I$.

The constraint requires $h^{(r)}$ not to go through the boundary of subsystem A. In the statistical-mechanical model of Ising spins, this corresponds to no domain wall going through the boundary of A, namely forcing $|\partial A|$ boundary spins aligning in the same direction (see Fig. 4). Thus, the negativity is associated with the excess free energy for aligning spins

$$\mathcal{E}_A^{(2n)} = \frac{1}{2n-2}\left(F_A^{(2n)} - F_0^{(2n)}\right) := \frac{\Delta F_A^{(2n)}}{2n-2}, \qquad \text{(A30)}$$

where $F_0^{(2n)} := -\log \mathcal{Z}_{2n,x}\mathcal{Z}_{2n,z}$ and $F_A^{(2n)}$ are the free energy without and with constraints, respectively. Since we have in total $2n-2$ constraints, $\mathcal{E}_A^{(2n)} = \Delta F_A$ with ΔF_A being the excess free energy for aligning one species of Ising spins.

APPENDIX B: $\mathbb{Z}_N$ TORIC CODE

So far, we focus only on the $\mathbb{Z}_2$ Toric code with incoherent errors. It is natural to inquire whether our methods are still applicable to $\mathbb{Z}_N$ Toric code and whether the results change. We provide a brief discussion on the $\mathbb{Z}_3$ Toric code in this subsection. We will use similar symbols to denote the basic operators and stabilizers, although their meanings are different from those in the $\mathbb{Z}_2$ case.

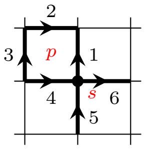
Let us first specify the Hamiltonian and the error models. Consider an $L \times L$ square lattice with periodic boundary conditions. The physical qutrits live on the edges of the lattice. We introduce the clock and shift operators

$$XZ = wZX, \quad w = e^{2\pi i/3},$$
$$Z = \begin{pmatrix} 1 & & \\ & w & \\ & & w^2 \end{pmatrix}, \quad X = \begin{pmatrix} & 1 & \\ & & 1 \\ 1 & & \end{pmatrix}. \tag{B1}$$

In and only in this subsection, X and Z refer to the clock and shift, respectively. The code subspace is given by the ground state subspace of the Hamiltonian

$$H_{\mathbb{Z}_3} = -\sum_s A_s - \sum_p B_p, \tag{B2}$$

where A_s and B_p are mutually commuting projectors associated with vertices and plaquettes, e.g.,



$$A_s = \frac{1}{3}\sum_{n=0}^{2}\left(X_4X_5X_1^{-1}X_6^{-1}\right)^n$$
$$B_p = \frac{1}{3}\sum_{n=0}^{2}\left(Z_4Z_1Z_2^{-1}Z_3^{-1}\right)^n \tag{B3}$$

One can verify that $A_s^2 = A_s$, $B_p^2 = B_p$. The ground state $|\Psi\rangle$ satisfies $A_s|\Psi\rangle = B_p|\Psi\rangle = |\Psi\rangle$, and the violation of A_s and B_p will be refered to as e (and its antiparticle $\bar{e}$) and m (and its antiparticle $\bar{m}$) anyons, respectively. For simplicity, we consider only the following incoherent error:

$$\mathcal{N}_{X,i}[\rho] = (1 - p_1 - p_2)\rho + p_1 Z_i \rho Z_i^\dagger + p_2 Z_i^2 \rho Z_i^{2,\dagger}, \tag{B4}$$

which creates a pair of e anyons in two different ways with probabilities p_1 and p_2. In the following, we will first assume $p_1 = p_2 = p$ and comment on what could change without this assumption.

To compute the three diagnostics, one can still work in the loop picture and map the nth momentum of the error-corrupted state to a partition function of a classical spin model that involves n-flavor three-state Potts spins. As the error rate increases, the spin model undergoes a paramagnet-to-ferromagnet transition. The three diagnostics are mapped to the corresponding observables in a similar fashion as what we have shown in the $\mathbb{Z}_2$ case. Therefore, they should undergo a transition simultaneously and yield a consistent characterization of the error-induced phase.

When $p_1 \neq p_2$, the spin models obtained in the loop picture contain complex phases and do not admit a statistical-mechanical interpretation. Technically, it brings sign problems to the Monte Carlo simulation. It is unclear whether the three diagnostics still exhibit transition simultaneously, which may be an interesting question for future study.

[1] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, Proceedings of Symposia in Applied Mathematics **68**, 13 (2010).

[2] A. R. Calderbank and P. W. Shor, Good quantum error correcting codes exist, Phys. Rev. A **54**, 1098 (1996).

[3] A. M. Steane, Error correcting codes in quantum theory, Phys. Rev. Lett. **77**, 793 (1996).

[4] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[5] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Non-Abelian anyons and topological quantum computation, Rev. Mod. Phys. **80**, 1083 (2008).

[6] X.-G. Wen, Colloquium: Zoo of quantum-topological phases of matter, Rev. Mod. Phys. **89**, eid 041004 (2017).

[7] A. Y. Kitaev, Fault tolerant quantum computation by anyons, Ann. Phys. **303**, 2 (2003).

[8] K. Fujii, Quantum computation with topological codes: From qubit to topological fault-tolerance, ArXiv:1504.01444 (2015).

[9] S. B. Bravyi and A. Y. Kitaev, Quantum codes on a lattice with boundary, ArXiv:9811052 (1998).

[10] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. **43**, 4452 (2002).

[11] D. Nigg, M. Müller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt, Quantum computations on a topologically encoded qubit, Science **345**, 302 (2014).

[12] K. J. Satzinger, *et al.*, Realizing topologically ordered states on a quantum processor, Science **374**, abi8378 (2021).

[13] R. Verresen, M. D. Lukin, and A. Vishwanath, Prediction of toric code topological order from Rydberg blockade, Phys. Rev. X **11**, 031005 (2021).

[14] G. Semeghini, *et al.*, Probing topological spin liquids on a programmable quantum simulator, Science **374**, abi8794 (2021).

[15] D. Bluvstein, *et al.*, A quantum processor based on coherent transport of entangled atom arrays, Nature **604**, 451 (2022).

[16] R. Acharya, *et al.*, Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, ArXiv:2207.06431 (2022).

[17] T. I. Andersen, *et al.*, Observation of non-Abelian exchange statistics on a superconducting processor, ArXiv:2210.10255 (2022).

[18] C. Wang, J. Harrington, and J. Preskill, Confinement Higgs transition in a disordered gauge theory and the accuracy threshold for quantum memory, Ann. Phys. **303**, 31 (2003).

[19] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, Error threshold for color codes and random three-body Ising models, Phys. Rev. Lett. **103**, 090501 (2009).

[20] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, Strong resilience of topological codes to depolarization, Phys. Rev. X **2**, 021004 (2012).

[21] A. Kubica, M. E. Beverland, F. Brandão, J. Preskill, and K. M. Svore, Three-dimensional color code thresholds via statistical-mechanical mapping, Phys. Rev. Lett. **120**, 180501 (2018).

[22] C. T. Chubb and S. T. Flammia, Statistical mechanical models for quantum codes with correlated noise, Ann. Inst. Henri Poincaré D **8**, 269 (2021).

[23] A. Y. Kitaev, A. Shen, M. N. Vyalyi, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Soc., 2002), Vol. 47.

[24] B. Schumacher and M. A. Nielsen, Quantum data processing and error correction, Phys. Rev. A **54**, 2629 (1996).

[25] B. Schumacher and M. D. Westmoreland, Approximate quantum error correction, Quantum Inf. Process. **1**, 5 (2002).

[26] M. Horodecki, J. Oppenheim, and A. Winter, Quantum state merging and negative information, Commun. Math. Phys. **269**, 107 (2006).

[27] T.-C. Lu, T. H. Hsieh, and T. Grover, Detecting topological order at finite temperature using entanglement negativity, Phys. Rev. Lett. **125**, 116801 (2020).

[28] T.-C. Lu and S. Vijay, Characterizing long-range entanglement in a mixed state through an emergent order on the entangling surface, ArXiv:2201.07792 (2022).

[29] M. B. Hastings, Topological order at nonzero temperature, Phys. Rev. Lett. **107**, eid 210501 (2011).

[30] E. H. Lieb and M. B. Ruskai, Proof of the strong subadditivity of quantum-mechanical entropy, J. Math. Phys. **14**, 1938 (1973).

[31] H. Araki, Relative entropy of states of von Neumann algebras, Publ. Res. Inst. Math. Sci. Kyoto **1976**, 809 (1976).

[32] G. Lindblad, Completely positive maps and entropy inequalities, Commun. Math. Phys. **40**, 147 (1975).

[33] K. Fredenhagen and M. Marcu, Charged states in Z2 gauge theories, Commun. Math. Phys. **92**, 81 (1983).

[34] K. Fredenhagen and M. Marcu, A confinement criterion for QCD with dynamical quarks, Phys. Rev. Lett. **56**, 223 (1986).

[35] K. Fredenhagen and M. Marcu, Dual interpretation of order parameters for lattice gauge theories with matter fields, Nucl. Phys. B Proc. Suppl. **4**, 352 (1988).

[36] K. Gregor, D. A. Huse, R. Moessner, and S. L. Sondhi, Diagnosing deconfinement and topological order, New J. Phys. **13**, 025009 (2011).

[37] L. Mazza, M. Rizzi, M. D. Lukin, and J. I. Cirac, Robustness of quantum memories based on Majorana zero modes, Phys. Rev. B **88**, 205142 (2013).

[38] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).

[39] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, Phys. Lett. A **223**, 1 (1996).

[40] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[41] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).

[42] M. B. Plenio, Logarithmic negativity: A full entanglement monotone that is not convex, Phys. Rev. Lett. **95**, 090503 (2005).

[43] X. Wen, S. Matsuura, and S. Ryu, Edge theory approach to topological entanglement entropy, mutual information and entanglement negativity in Chern-Simons theories, Phys. Rev. B **93**, 245140 (2016).

[44] X. Wen, P.-Y. Chang, and S. Ryu, Topological entanglement negativity in Chern-Simons theories, JHEP **09**, 012 (2016).

[45] H. Shapourian, K. Shiozaki, and S. Ryu, Partial time-reversal transformation and entanglement negativity in fermionic systems, Phys. Rev. B **95**, 165101 (2017).

[46] A. Kitaev and J. Preskill, Topological entanglement entropy, Phys. Rev. Lett. **96**, 110404 (2006).

[47] M. Levin and X.-G. Wen, Detecting topological order in a ground state wave function, Phys. Rev. Lett. **96**, eid 110405 (2006).

[48] T. Grover, A. M. Turner, and A. Vishwanath, Entanglement entropy of gapped phases and topological order in three dimensions, Phys. Rev. B **84**, 195120 (2011).

[49] R. B. Potts, Some generalized order-disorder transformations, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 48 (Cambridge University Press, 1952), p. 106.

[50] T. Kihara, Y. Midzuno, and T. Shizume, Statistics of two-dimensional lattices with many components, J. Phys. Soc. Jpn. **9**, 681 (1954).

[51] M. Kohmoto, M. den Nijs, and L. P. Kadanoff, Hamiltonian studies of the $d = 2$ Ashkin-Teller model, Phys. Rev. B **24**, 5229 (1981).

[52] K. Binder and D. P. Landau, Finite-size scaling at first-order phase transitions, Phys. Rev. B **30**, 1477 (1984).

[53] S. Iino, S. Morita, N. Kawashima, and A. W. Sandvik, Detecting signals of weakly first-order phase transitions in two-dimensional Potts models, J. Phys. Soc. Jpn. **88**, 034006 (2019).

[54] M. A. Metlitski, C. A. Fuertes, and S. Sachdev, Entanglement entropy in the O(N) model, Phys. Rev. B **80**, 115122 (2009).

[55] We thank Tarun Grover for pointing it out to us.

[56] Obtaining the negativity from the Monte Carlo simulation is not an easy task. Here, one directly computes $e^{(2-n)\mathcal{E}_A^{(n)}}$, which is exponentially small due to the area-law scaling of $\mathcal{E}_A^{(n)}$ and thus requires exponentially many samples to accurately determine its value. This limits the largest accessible subsystem size.

[57] H.-C. Jiang, R. R. P. Singh, and L. Balents, Accuracy of topological entanglement entropy on finite cylinders, Phys. Rev. Lett. **111**, 107205 (2013).

[58] H. Nishimori, Internal energy: Specific heat and correlation function of the bond-random ising model, Prog. Theor. Phys. **66**, 1169 (1981).

[59] H. Nishimori and K. Nemoto, Duality and multicritical point of two-dimensional spin glasses, J. Phys. Soc. Jpn. **71**, 1198 (2002).

[60] M. Ohzeki, Spin glass a bridge between quantum computation and statistical mechanics, in *Lectures on Quantum Computing, Thermodynamics and Statistical Physics* (2013), p. 63.

[61] One can show the usual Kramers-Wannier duality between the two spin models by performing high- and low-temperature expansions.

[62] A. Honecker, M. Picco, and P. Pujol, Universality class of the Nishimori point in the 2d±j random-bond Ising model, Phys. Rev. Lett. **87**, 047201 (2001).

[63] We note that the replicated RBIM in the limit $n \to 1$ contains a summation over different boundary conditions. However, the critical point does not depend on the

boundary condition as the ferromagnetic transition is a bulk transition.

[64] A. S. Darmawan and D. Poulin, Tensor-network simulations of the surface code under realistic noise, Phys. Rev. Lett. **119**, eid 040502 (2017).

[65] A. S. Darmawan and D. Poulin, Linear-time general decoding algorithm for the surface code, Phys. Rev. E **97**, eid 051302 (2018).

[66] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, npj Quantum Inf. **4**, eid 55 (2018).

[67] F. Venn, J. Behrends, and B. Béri, Coherent error threshold for surface codes from Majorana delocalization, ArXiv:2211.00655 (2022).

[68] C. G. Brell, S. Burton, G. Dauphinais, S. T. Flammia, and D. Poulin, Thermalization, error-correction, and memory lifetime for Ising anyon systems, Phys. Rev. X **4**, 031058 (2014).

[69] J. R. Wootton, J. Burri, S. Iblisdir, and D. Loss, Error correction for non-Abelian topological quantum computation, Phys. Rev. X **4**, 011051 (2014).

[70] A. Schotte, G. Zhu, L. Burgelman, and F. Verstraete, Quantum error correction thresholds for the universal Fibonacci Turaev-Viro code, Phys. Rev. X **12**, 021012 (2022).

[71] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, Topological order and decoherence-induced transitions in mixed states, (to be published).

[72] H. Song, J. Schönmeier-Kromer, K. Liu, O. Viyuela, L. Pollet, and M. A. Martin-Delgado, Optimal thresholds for Fracton Codes and Random Spin Models with Subsystem Symmetry, Phys. Rev. Lett. **129**, 230502 (2022).

[73] P. Le Doussal and A. B. Harris, Location of the Ising spin-glass multicritical point on Nishimori's line, Phys. Rev. Lett. **61**, 625 (1988).

[74] I. A. Gruzberg, N. Read, and A. W. Ludwig, Random-bond Ising model in two dimensions: The Nishimori line and supersymmetry, Phys. Rev. B **63**, eid 104422 (2001).

[75] J. Y. Lee, C.-M. Jian, and C. Xu, Quantum criticality under decoherence or weak-measurement (to be published).