Information dynamics in decohered quantum memory with repeated syndrome measurements: a dual approach

Jacob Hauser,[1] Yimu Bao,[2] Shengqi Sang,[2, 3, 4] Ali Lavasani,[2] Utkarsh Agrawal,[2] and Matthew P. A. Fisher[1]

[1]*Department of Physics, University of California, Santa Barbara, CA 93106, USA*
[2]*Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA*
[3]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[4]*Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

Measurements can detect errors in a decohered quantum memory allowing active error correction to increase the memory time. Previous understanding of this mechanism has focused on evaluating the performance of error correction algorithms based on measurement results. In this work, we instead intrinsically characterize the information dynamics in a quantum memory under repeated measurements, using coherent information and relative entropy. We consider the dynamics of a d-dimensional stabilizer code subject to Pauli errors and noisy stabilizer measurements and develop a $(d + 1)$-dimensional statistical mechanics model for the information-theoretic diagnostics. Our model is dual to the model previously obtained for the optimal decoding algorithm, and the potential decoding transition in the quantum memory again manifests as a thermal phase transition in the statistical mechanics model. We explicitly derive the model and study the phase transition in information encoding in three examples: surface codes, repetition codes, and the XZZX code.

I. INTRODUCTION

Protecting quantum information against decoherence in realistic devices is a central focus of quantum information science. While designing a passive memory that encodes quantum information in thermal equilibrium states remains elusive below four dimensions, one can alternatively perform active error correction to protect encoded information in non-equilibrium states [1, 2]. This entails using repeated measurements to detect error syndromes and applying feedback based on the measurement results to correct errors. Measurements remove entropy from the system and therefore can compete with decoherence to increase the memory time, even if the measurement outcomes are noisy.

Prior work has sought to understand this mechanism by examining the performance of specific decoding algorithms for specific quantum error-correcting (QEC) codes. For example, in a surface code subject to decoherence, repeated noisy syndrome measurements enable a decoding algorithm to protect quantum information until exponentially late times as long as the rates of decoherence and measurement noise are below a finite threshold [3]. This is demonstrated in Ref. [3] by mapping the transition in the fidelity of the decoding algorithm to the confinement transition in the 3D random-plaquette Ising model (RPIM), i.e. a random $\mathbb{Z}_2$ gauge theory. It is natural to ask whether the impact of repeated syndrome measurements on memory time can be understood intrinsically from the information dynamics of the system without reference to a particular decoding algorithm.

It has been recently shown that the surface code with local decoherence can undergo an intrinsic information transition in the mixed density matrix of the system, which governs the decoding transition based on a single round of perfect measurements [4–13]. In particular, Ref. [4] maps the coherent information (the amount of re-

coverable information from the mixed state), along with several other information diagnostics, to observables in a 2D multi-flavor Ising model. At a finite decoherence rate, the multi-flavor Ising model undergoes a ferromagnetic transition, detected by the information diagnostics. Notably, the multi-flavor Ising model is dual to the random-bond Ising model along the Nishimori line [14], which characterizes the decoding transition for the optimal decoding algorithm [3]. This duality stems from two different expansions of the same decohered density matrix [4].

In this work, we extend this perspective to general stabilizer codes subject to local decoherence and repeated imperfect syndrome measurements. We study information dynamics characterized by intrinsic information diagnostics, including coherent information and relative entropy. To this end, we map the Rényi versions of the diagnostics to observables in statistical mechanics (stat-mech) models based on the stabilizer expansion of the physical density matrix. We further perform the expansion of the same density matrix in terms of error configurations and obtain the stat-mech model previously derived for the optimal decoding algorithm (i.e. the maximum likelihood decoder) [3, 15–22]. The two stat-mech models are related by high-to-low temperature duality. In the example of decohered surface codes under repeated syndrome measurements, the decoding transition manifests as a ferromagnetic transition in a 3D multi-flavor Ising model (in the stabilizer expansion) dual to the confinement transition in the 3D RPIM (in the error configuration expansion).

A. Overview

We begin with an overview of our results before delving into the details. We consider the dynamics of d-dimensional stabilizer codes subject to local Pauli errors and repeated stabilizer measurements. The dynamics
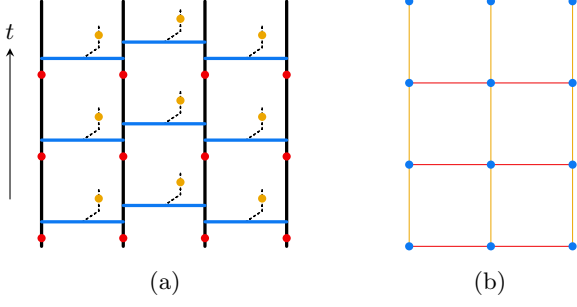
Figure 1. (a) One-dimensional repetition code under repeated measurements. Each black solid line denotes a physical qubit. In every time step, check operators $Z_r Z_{r+1}$ on neighboring qubits are measured (horizontal blue lines) with measurements modeled by coupling measurement ancillae (black dashed lines) to physical qubits. The red and orange circles denote the physical decoherence and measurement errors on the physical qubits and ancillae, respectively. (b) The 2D multi-flavor Ising model for the 1D repetition code. The spins live on the vertices (blue dots) of a 2D square lattice; each vertex corresponds to a check operator in the circuit. The spatial couplings (in red) are related to the strength of physical decoherence. The temporal couplings (in orange) are related to the error rate of measurement readout.

is described by a quantum circuit of T steps as illustrated in Fig. 1(a). In each time step, we first apply single-qubit Pauli channels to the physical qubits and then perform noisy measurements on a set of stabilizers (called check operators). To characterize the information dynamics, we consider the coherent information, the amount of recoverable information in the system together with the measurement record, and the relative entropy between the quantum states of the memory for two initial states. The key result of this work is to develop a $(d+1)$-dimensional stat-mech model for the n-th moment $\operatorname{tr} \rho_{Q\mathbf{M}}^n$ of the density matrix of the system Q together with measurement record $\mathbf{M}$ as illustrated in Fig. 1(b). The mapping allows evaluating the Rényi versions of our information diagnostics as observables in the stat-mech model.

In Sec. IV A, we detail the mapping from the n-th moment $\operatorname{tr} \rho_{Q\mathbf{M}}^n$ to the partition function of the stat-mech model based on the stabilizer expansion of $\rho_{Q\mathbf{M}}$. The model involves $n - 1$ flavors of Ising spins and has a Hamiltonian

$$\mathcal{H}^{(n)}(\{\boldsymbol{\sigma}^{\mathsf{a}}\}) = \sum_{\mathsf{a}} \mathcal{H}(\boldsymbol{\sigma}^{\mathsf{a}}) + \mathcal{H}(\textstyle\prod_{\mathsf{a}} \boldsymbol{\sigma}^{\mathsf{a}}), \qquad (1)$$

where $\boldsymbol{\sigma}^{\mathsf{a}}$ denotes a collection of the a-th flavor of Ising spins. The Hamiltonian $\mathcal{H}(\boldsymbol{\sigma}^{\mathsf{a}})$ for each flavor involves $T + 1$ layers of Ising spins for T time steps. Each spin $\sigma_{t,i}$ in the t-th layer is associated with a check operator g_i. The Hamiltonian $\mathcal{H}(\boldsymbol{\sigma})$ contains intralayer coupling $H(\boldsymbol{\sigma}_t)$ determined by the decoherence rate and interlayer Ising couplings with strength $J_q = -(1/2)\log(1 - 2q)$

controlled by the measurement error rate q, i.e.

$$\mathcal{H}(\boldsymbol{\sigma}) = \sum_{t=0}^{T-1} H(\boldsymbol{\sigma}_t) - J_q \sum_{t=0}^{T-1} \sum_i \sigma_{t,i}\sigma_{t+1,i}, \qquad (2)$$

where we use $\boldsymbol{\sigma}_t$ and $\boldsymbol{\sigma} := \{\boldsymbol{\sigma}_t\}$ to denote the spin configuration at time t and the entire spacetime, respectively. We note that in this model, higher noise rates give rise to larger couplings (equiv. lower temperatures).

The multi-flavor Ising model exhibits a symmetry that depends on the redundancy in the set of measured check operators. A subset of check operators is redundant if its product is the identity, and such redundancies form a group $\mathcal{R}$. The symmetry of the multi-flavor Ising model is then given by $\mathbb{G}^{(n)} = (\mathcal{R}^n \rtimes S_n)/\mathcal{R}$. Notably, $\mathbb{G}^{(n)}$ is either a local or global symmetry depending on whether the redundant subset of check operators has local or global support. For example, in the 2D toric code under repeated measurements of plaquette and star operators, the redundancy $\mathcal{R} = \mathbb{Z}_2 \times \mathbb{Z}_2$ is given by the product of all plaquette operators and the product of all star operators, resulting in a global symmetry of the stat-mech model. In the 2D repetition code on a square lattice with check operators on the edges, the redundancy is given by the product of four check operators locally, giving rise to a local (gauge) symmetry of the stat-mech model. When tuning the decoherence rate, the stat-mech model can undergo a phase transition that either spontaneously breaks the global symmetry $\mathbb{G}^{(n)}$ or confines the gauge field.

The information diagnostics map to distinct observables in the stat-mech model. Specifically, the coherent information is associated with the excess free energy of inserting defects in the stat-mech model, and the relative entropy maps to correlation functions. In the examples studied in this work (surface codes, repetition codes, and the XZZX code), such observables detect the underlying phase transition in the stat-mech model and exhibit singularities at the same point.

We note that the boundary condition in the stat-mech model depends on the diagnostic we consider. For the information diagnostics associated with both the system Q and the measurement record $\mathbf{M}$, the stat-mech model has an open boundary condition. In contrast, the stat-mech model for the diagnostics associated with only the measurement record has an external field at the final layer of the spins.

The multi-flavor model obtained in the stabilizer expansion is dual to the stat-mech model previously derived for the maximum likelihood (ML) decoder. In Sec. IV B, we explicitly map the n-th moment $\operatorname{tr} \rho_{Q\mathbf{M}}^n$ to a stat-mech model based on the error configuration expansion. Such a model reduces to that for the ML decoder in the limit $n \to 1$ and is the Kramers-Wannier dual of the multi-flavor Ising model.

Before proceeding, we comment on the advantages of the multi-flavor Ising model obtained in the stabilizer expansion. First, the mapping excels at identifying codes with maximum threshold. In our model, the

high-temperature phase corresponds to the low error rate regime of the stabilizer code. Thus, a code with a maximum error threshold (e.g. $p_c = 0.5$ in the toric code with Pauli-Y error) corresponds to a stat-mech model with no finite-temperature ordering. These models are better understood compared to their dual (obtained previously for the ML decoder), which are highly constrained and order up to infinite temperature. Second, the impact of measurement error is clear in the multi-flavor Ising model. A higher rate of measurement readout errors corresponds to a stronger interlayer coupling, which favors the ferromagnetic phase that does not encode information. Third, the benefit of repeating noisy syndrome measurements in quantum error correction is also explicit in our mapping. Successful error correction relies on using the measurement record to identify the error configurations in the system. The corresponding stat-mech model has an external field on the final layer of spins due to the measurement readout errors. When considering QEC in the decohered 2D surface code, the presence of this symmetry-breaking field indicates the decodable phase (i.e. a paramagnetic phase) can exist only when the measurements are repeated an extensive number of times.

The rest of the paper is organized as follows. Sec. II describes the dynamics of stabilizer code under repeated noisy syndrome measurements. Sec. III introduces the diagnostics of information dynamics in quantum memory. Sec. IV develops stat-mech models for information diagnostics based on two expansions of the density matrix. We close with discussion in Sec. V.

CONTENTS

II. SETUP

In this work, we study the dynamics of quantum information in stabilizer codes under decoherence and repeated noisy syndrome measurements. We keep track of the mixed density matrix describing the quantum state of the system together with the classical measurement outcomes. In what follows, we first review the definition of stabilizer codes and then detail the dynamics considered in this work.

Stabilizer codes are a broad class of QEC codes whose code space $\mathcal{L}$ is the simultaneous $+1$ eigenspace of a set of mutually commuting Pauli operators g_i [23]

$$\mathcal{L} = \text{span}\{|\psi\rangle\langle\psi| : \mathsf{g}_i|\psi\rangle = |\psi\rangle \,,\, i \in 1, 2, \ldots, I_0\}. \quad (3)$$

The Pauli operators generate a stabilizer group under multiplication[1]

$$\mathcal{S} = \langle \mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_{I_0} \rangle. \quad (4)$$

Each element $s \in \mathcal{S}$ is called a *stabilizer* and takes a $+1$ eigenvalue in the code space. We take the I_0 generators to be independent; the stabilizer group then contains 2^{I_0} elements and is isomorphic to $\mathbb{Z}_2^{I_0}$.

A stabilizer code with N physical qubits can encode $K = N - I_0$ logical qubits. We use $\bar{g}_k^x$, $\bar{g}_k^z$ for $k = 1, 2, \ldots, K$ to denote the Pauli-X and -Z operators of logical qubits. These operators commute with the stabilizers and operate within the code space. In practice, stabilizer measurements are performed to identify errors in a stabilizer code; $s|\psi\rangle = -|\psi\rangle$ for any $s \in \mathcal{S}$ indicates a departure from the code space. We refer to the collection of stabilizer measurement outcomes as an *error syndrome*.

We are now ready to introduce the stabilizer code dynamics studied in this work. We consider a stabilizer code with N physical qubits initialized in an encoded state $\rho_0 \in \mathcal{L}$ and evolved under decoherent interaction with the environment interrupted by repeated syndrome measurements (as shown in Fig. 1) for T time steps. In particular, at each time step $t \in \{1, \ldots, T\}$ the following occurs:

1. We apply single-qubit Pauli channels to every qubit in the system Q, with p_x, p_y, and p_z denoting the probability of each type of Pauli error[2].

2. We measure a set of stabilizers $\mathcal{C}$ and record their measurement results in the ancilla qubits in M_t. Each measurement may have a readout error; its result is recorded incorrectly with probability q.

The decoherence at each time step is described by $\mathcal{N}[\rho] = \prod_r \mathcal{N}_r[\rho]$ where the single-qubit Pauli channel takes the form

$$\mathcal{N}_r[\rho] = \left(1 - \sum_\alpha p_\alpha\right)\rho + \sum_\alpha p_\alpha P_r^\alpha \rho P_r^\alpha. \quad (5)$$

Here, $P_r^\alpha = X_r, Y_r, Z_r$ for $\alpha = x, y, z$, and their corresponding error rates satisfy $p_\alpha > 0$ and $\sum_\alpha p_\alpha \leq 1$. In

this work, we focus on the regime $p_{\alpha_1} + p_{\alpha_2} \leq 1/2$ for any $\alpha_1 \neq \alpha_2$.

We perform measurements on the same set of check operators in the stabilizer group at each time step,

$$\mathcal{C} := \{g_1, g_2, \ldots, g_I\}, \quad (6)$$

where each $g_i \in \mathcal{S}$. We focus on the case that $\mathcal{C}$ generates the entire stabilizer group. However, the check operators in $\mathcal{C}$ may form an overcomplete generating set with redundant check operators given by a product of the others. The redundancies in check operators lead to constraints in measurement results in the absence of measurement errors and may provide additional information when performing error correction with noisy measurement results. These redundancies form a group $\mathcal{R}$ that is isomorphic to $\mathbb{Z}_2^{I-I_0}$. In the 1D repetition code with a periodic boundary condition, the product of all check operators is the identity, giving rise to a global $\mathbb{Z}_2$ redundancy. In the 2D repetition code on the square lattice, the product of check operators around a plaquette is the identity, giving rise to local $\mathbb{Z}_2$ redundancies.

A measurement can be modeled by coupling the system Q to an ancilla M via a unitary operator and then projectively measuring the ancilla. Specifically, the unitary operator acting on the system and ancilla prepared in $|0\rangle_M$ defines an isometric embedding

$$U_{Q \to QM} = \mathcal{P}_0 \otimes |0\rangle_M + \mathcal{P}_1 \otimes |1\rangle_M, \quad (7)$$

where $\mathcal{P}_{0,1} = (1 \pm g)/2$ is the projector associated with the outcome ± 1 when measuring a check operator g, and $(U_{Q \to QM})^\dagger U_{Q \to QM} = \mathbb{1}_Q$. A subsequent measurement on ancilla M in the Pauli-Z basis reads out the outcome and projects the system Q to the corresponding eigenstate of the check operator g. This measurement outcome is probabilistic with its probability determined by the Born rule, $p_{0,1} = \text{tr}\,\mathcal{P}_{0,1}\rho$. We note that the results in the paper also apply to the case of weak measurements, as discussed in Appendix A.

Measurements in realistic quantum devices also have readout errors modeled by quantum channels acting on the ancillary measurement device. In our dynamics, such errors occur with probability q and can be described by a bit-flip channel before measuring the ancilla:

$$\mathcal{N}_M[\rho_M] = (1 - q)\rho_M + q X_M \rho_M X_M. \quad (8)$$

Here, instead of considering individual states after measurements, we keep track of the ensemble (i.e. an incoherent mixture) of the post-measurement states on the system and ancilla, which can be obtained by applying complete dephasing channels on the ancillae in the measurement basis. This ensemble contains, on average, the recoverable information. However, any code state subject to Pauli errors and strong syndrome measurements always remains block-diagonal in the measurement basis; the complete dephasing does not change the density matrix at all and therefore can be neglected. The ensemble

---

[1] Without loss of generality, we set the phase of each Pauli operator to be $+1$.

[2] Although we focus on uncorrelated Pauli errors in this work, our framework can be straightforwardly extended to correlated Pauli errors.

generated by noisy measurements is then given by the quantum channel

$$\mathcal{M}[\rho] = \mathcal{N}_M[U_{Q \to QM} \, \rho \, U_{Q \to QM}^\dagger], \qquad (9)$$

which we refer to as the noisy measurement channel.

Having specified the quantum memory dynamics, we can write down the evolution of the density matrix

$$\begin{aligned}
\rho_{Q\mathbf{M}_t} &= \mathcal{M}_t \circ \mathcal{N}[\rho_{Q\mathbf{M}_{t-1}}] \\
&= \sum_{\boldsymbol{m}_t} \mathcal{P}_{\boldsymbol{m}_t} \mathcal{N}[\rho_{Q\mathbf{M}_{t-1}}] \mathcal{P}_{\boldsymbol{m}_t} \otimes \mathcal{N}_{M_t} \left[|\boldsymbol{m}_t\rangle\!\langle\boldsymbol{m}_t| \right],
\end{aligned}$$
$$(10)$$

where $\mathcal{M}_t$ is the measurement channel for measuring all the stabilizers at time t, $\mathcal{N}_{M_t}$ is the measurement error channel acting on the ancillae introduced at time t, and $\mathcal{P}_{\boldsymbol{m}_t}$ is the projector associated with the collection of measurement outcomes $\boldsymbol{m}_t$. Here, we use $\mathbf{M}_t = \{M_1, M_2, \ldots, M_t\}$ to denote all the ancilla qubits introduced up to step t.

Before proceeding, we introduce a notation for the Pauli operator $\mathcal{O}_{\boldsymbol{a}}$ in terms of a $2N$-component binary vector $\boldsymbol{a} = (\boldsymbol{a}^x, \boldsymbol{a}^z)$,

$$\mathcal{O}_{\boldsymbol{a}} = \mathrm{i}^{\boldsymbol{a}^x \cdot \boldsymbol{a}^z} X^{\boldsymbol{a}^x} Z^{\boldsymbol{a}^z} := \mathrm{i}^{\boldsymbol{a}^x \cdot \boldsymbol{a}^z} \prod_{r=1}^{N} X_r^{a_r^x} Z_r^{a_r^z}. \qquad (11)$$

In this way, we can label the stabilizers using the binary vectors. We introduce binary vectors $\boldsymbol{a}_i$ to label the check operators $g_i = \mathcal{O}_{\boldsymbol{a}_i}$. The set of vectors $\{\boldsymbol{a}_i\}$ for check operators unambiguously specifies the stabilizer code. Then, an arbitrary stabilizer $g^{\boldsymbol{u}} \in \mathcal{S}$ may be denoted using an I-component binary vector $\boldsymbol{u}$,

$$g^{\boldsymbol{u}} := \prod_{i=1}^{I} g_i^{u_i}. \qquad (12)$$

If the set of check operators $\mathcal{C}$ is overcomplete, the redundancies are labeled by $g^{\boldsymbol{u}} = \mathbb{1}$ for $\boldsymbol{u} \neq 0$, and the redundancy group is $\mathcal{R} = \{\boldsymbol{u} : g^{\boldsymbol{u}} = \mathbb{1}\}$.

We also introduce binary vectors $\overline{\boldsymbol{a}}_k$ for $k = 1, \ldots, K$ ($k = K+1, \ldots, 2K$) to label the logical Pauli-X(Z) operator $\overline{g}_k^x = \mathcal{O}_{\overline{\boldsymbol{a}}_k}$ ($\overline{g}_k^z = \mathcal{O}_{\overline{\boldsymbol{a}}_{k+K}}$). We can denote any Pauli operator in the logical space using a $2K$-component binary vector $\boldsymbol{\kappa}$,

$$\overline{g}^{\boldsymbol{\kappa}} := \prod_{k=1}^{K} \mathrm{i}^{\kappa_k \kappa_{k+K}} (\overline{g}_k^x)^{\kappa_k} (\overline{g}_k^z)^{\kappa_{k+K}}. \qquad (13)$$

III. INFORMATION DIAGNOSTICS

In this section, we introduce coherent information (Sec. III A) and relative entropy (Sec. III B) as diagnostics of information dynamics in quantum memories. In Sec. III C, we discuss the relation between coherent information and the fidelity of the optimal decoder.

A. Coherent information

A central quantity to characterize information dynamics is the *coherent information*, which measures the amount of recoverable information in the quantum memory together with the measurement outcomes [24–27].

To formulate this quantity, one can consider the stabilizer code initialized in a maximally entangled Bell state with a reference R. The amount of recoverable information in the system Q together with the measurement record $\mathbf{M} := \mathbf{M}_T$ is given by the recoverable entanglement with the reference R,

$$I_c(R\rangle Q\mathbf{M}) := S(\rho_{Q\mathbf{M}}) - S(\rho_{Q\mathbf{M}R}), \qquad (14)$$

where $S(\rho_{Q\mathbf{M}(R)}) := -\operatorname{tr} \rho_{Q\mathbf{M}(R)} \log \rho_{Q\mathbf{M}(R)}$ is the von Neumann entropy. We note that the coherent information retained in an ensemble of quantum trajectories (i.e. quantum states labeled by the history of measurement outcomes) is also studied in the context of measurement-induced transitions [28].

Analytically computing coherent information is challenging. Here, we consider the Rényi version of the coherent information

$$I_c^{(n)}(R\rangle Q\mathbf{M}) = \frac{1}{1-n} \log \frac{\operatorname{tr} \rho_{Q\mathbf{M}}^n}{\operatorname{tr} \rho_{Q\mathbf{M}R}^n}, \qquad (15)$$

which recovers the coherent information in the limit $n \to 1$.

We remark that the Rényi version of the coherent information in Eq. (15) is different from the average Rényi coherent information over measurement trajectories. Specifically, one can express $I_c^{(n)}$ in terms of the n-th moment in each trajectory $\boldsymbol{m}$ as

$$I_c^{(n)}(R\rangle Q\mathbf{M}) = \frac{1}{1-n} \log \frac{\sum_{\boldsymbol{m}} p_{\boldsymbol{m}}^n \operatorname{tr} \rho_{Q,\boldsymbol{m}}^n}{\sum_{\boldsymbol{m}} p_{\boldsymbol{m}}^n \operatorname{tr} \rho_{QR,\boldsymbol{m}}^n}. \qquad (16)$$

Since the averages over $\boldsymbol{m}$ occur inside the logarithm and are weighted by the n-th power of the trajectory probability $p_{\boldsymbol{m}}$, this quantity is distinct from the average of the Rényi coherent information over trajectories. Across a decoding transition in quantum memory (e.g. as in Appendix D), we expect the two quantities to exhibit different critical points and universalities for $n > 1$. The two quantities coincide only when $n \to 1$.

B. Relative entropy

Intuitively, retaining encoded information requires the final state of the stabilizer code to depend on its initial state. We thus consider initializing the stabilizer code in distinct initial states and use the distinguishability between their corresponding final states, measured by the *relative entropy*, to characterize the information flow.

First, we consider the quantum relative entropy between the density matrices of the final states of the system Q together with the measurement record $\mathbf{M}$. Here,

we prepare the stabilizer code in two orthogonal states, a logical state $|\Psi\rangle$ and an excited state $|\Psi_{\boldsymbol{s}}\rangle = w_{\boldsymbol{s}}|\Psi\rangle$, where $w_{\boldsymbol{s}}$ is an error operator that creates the syndrome $\boldsymbol{s} \in \mathbb{Z}_2^I$[3]. The relative entropy of the two corresponding final states $\rho_{Q\mathbf{M}}$ and $\rho_{Q\mathbf{M},\boldsymbol{s}}$ is given by

$$D_{\boldsymbol{s}} := \operatorname{tr} \rho_{Q\mathbf{M}} \log \rho_{Q\mathbf{M}} - \operatorname{tr} \rho_{Q\mathbf{M}} \log \rho_{Q\mathbf{M},\boldsymbol{s}}. \quad (17)$$

To compute the relative entropy, one can again formulate it as the $n \to 1$ limit of the Rényi relative entropies

$$D_{\boldsymbol{s}}^{(n)} = \frac{1}{n-1} \log \frac{\operatorname{tr} \rho_{Q\mathbf{M}}^n}{\operatorname{tr} \rho_{Q\mathbf{M}} \rho_{Q\mathbf{M},\boldsymbol{s}}^{n-1}}. \quad (18)$$

Alternatively, one can consider the relative entropy between two density matrices of the measurement record $\mathbf{M}$ only. This concerns whether the measurement record can correctly identify the syndromes, which is required for successful quantum error correction. The density matrix $\rho_{\mathbf{M}}$ is diagonal and represents the probability distribution of measurement outcomes, i.e. $\rho_{\mathbf{M}} = \sum_{\boldsymbol{m}} p_{\boldsymbol{m}} |\boldsymbol{m}\rangle\langle\boldsymbol{m}|$. The relative entropy then reduces to the Kullback–Leibler (KL) divergence between two distributions

$$D_{\mathrm{KL},\boldsymbol{s}} := \sum_{\boldsymbol{m}} p_{\boldsymbol{m}} \log p_{\boldsymbol{m}} - p_{\boldsymbol{m}} \log p_{\boldsymbol{m},\boldsymbol{s}}, \quad (19)$$

where $p_{\boldsymbol{m},\boldsymbol{s}}$ denotes the probability of trajectory $\boldsymbol{m}$ when initializing the code in the excited state $|\Psi_{\boldsymbol{s}}\rangle$. The KL divergence can be similarly expressed as the $n \to 1$ limit of the Rényi sequence

$$D_{\mathrm{KL},\boldsymbol{s}}^{(n)} := \frac{1}{n-1} \log \frac{\sum_{\boldsymbol{m}} p_{\boldsymbol{m}}^n}{\sum_{\boldsymbol{m}} p_{\boldsymbol{m}} p_{\boldsymbol{m},\boldsymbol{s}}^{n-1}}. \quad (20)$$

We note that both the quantum relative entropy $D_{\boldsymbol{s}}$ and the KL divergence $D_{\mathrm{KL},\boldsymbol{s}}$ are monotones under completely positive trace-preserving maps [29–32], and therefore can only decrease in the dynamics considered in this work. As we show in Sec. IV A, the two quantities map to correlation functions in the stat-mech model with different boundary conditions and can detect the potential transition corresponding to the information loss as demonstrated in Appendix D. Both quantities have been studied in many-body settings; the quantum relative entropy is used to probe the decoding transitions in the decohered toric code [4, 6], and the KL divergence between measurement records is used to diagnose measurement-induced transitions [33].

—————

[3] The logical (the excited) state need not be a pure state and can be also a mixed state describing the incoherent superposition of different logical states (excited states with the same syndrome $\boldsymbol{s}$).

C. Fidelity of the optimal decoder

Previous works have characterized the information dynamics in quantum memories using the performance of decoding algorithms. In particular, extensive work has been done to compute the fidelity of the *maximum likelihood* (ML) decoder for various decohered quantum codes and to use this as a benchmark of the information encoding, both with and without errors in syndrome measurements [3, 17, 19, 34, 35]. In what follows, we first briefly review the ML decoder and then demonstrate that the ML decoder is asymptotically optimal for Pauli errors by establishing a relation between its fidelity and coherent information.

To evaluate the decoding fidelity, it is convenient to expand the density matrix $\rho_{Q\mathbf{M}}$ as an incoherent superposition of states with different error configurations

$$\rho_{Q\mathbf{M}} = \sum_{\{\boldsymbol{b}_t, \boldsymbol{\epsilon}_t\}} \Pr(\{\boldsymbol{b}_t, \boldsymbol{\epsilon}_t\}) \mathcal{O}_{\Sigma\boldsymbol{b}_T} \rho_0 \mathcal{O}_{\Sigma\boldsymbol{b}_T}$$
$$\otimes_t |\boldsymbol{m}_t + \boldsymbol{\epsilon}_t\rangle\langle\boldsymbol{m}_t + \boldsymbol{\epsilon}_t|. \quad (21)$$

where $\boldsymbol{b}_t = (\boldsymbol{b}_t^x, \boldsymbol{b}_t^z)$ denotes the Pauli error applied on Q, the binary vector $\boldsymbol{\epsilon}_t$ denotes syndrome measurements that are erroneously flipped at time step t, and $\Sigma\boldsymbol{b}_t := \sum_{\tau=1}^t \boldsymbol{b}_\tau$ is the cumulative Pauli error. Here, we can omit the projector $\mathcal{P}_{\boldsymbol{m}_t}$ at each step in Eq. (10) because the syndrome $\boldsymbol{m}_t$ is completely determined by the accumulated Pauli error $\mathcal{O}_{\Sigma\boldsymbol{b}_T}$. The probability $\Pr(\{\boldsymbol{b}_t, \boldsymbol{\epsilon}_t\})$ for a given error configuration is a positive number determined by the error rates p_α and q.

In a decoding algorithm, our goal is to remove the excitations created by the cumulative error operator $\mathcal{O}_{\Sigma\boldsymbol{b}_T}$ based on the noisy measurement record $\{\boldsymbol{m}_t' := \boldsymbol{m}_t + \boldsymbol{\epsilon}_t\}$. Here, we assume that the last round of measurement is perfect, i.e. $\boldsymbol{\epsilon}_T = 0$, and the record contains the genuine measurement outcome $\boldsymbol{m}_T$ of the check operators in the final state. In this case, one can restore a logical state by applying a recovery Pauli operator to completely remove the excitations. However, the recovery operator is not unique, and applying a recovery operator that differs from the cumulative error by a logical operator $\overline{g}^{\boldsymbol{\kappa}}$ may result in a logical error. In the maximum likelihood decoder, one evaluates the probability $\Pr(\boldsymbol{\kappa}|\boldsymbol{m}_T, \{\boldsymbol{m}_t'\})$ of cumulative error operators in the class labeled by $\boldsymbol{\kappa}$ and selects the recovery operator in the class $\boldsymbol{\kappa}^{\mathrm{ML}}$ with the maximum probability among 2^{2K} classes

$$\boldsymbol{\kappa}^{\mathrm{ML}}(\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}) := \operatorname{argmax}_{\boldsymbol{\kappa}} \Pr(\boldsymbol{\kappa}|\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}). \quad (22)$$

It follows that the ML decoder's decoding failure probability for a given syndrome is

$$\Delta(\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}) := 1 - \max_{\boldsymbol{\kappa}} \Pr(\boldsymbol{\kappa}|\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}). \quad (23)$$

In Appendix B, we show that the average failure probability $\overline{\Delta} := \overline{\Delta(\boldsymbol{m}_T, \{\boldsymbol{m}_t'\})}$ has an upper bound:

$$\overline{\Delta} \le H(\boldsymbol{\kappa}|\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}), \quad (24)$$

where $H(\boldsymbol{\kappa}|\boldsymbol{m}_T, \{\boldsymbol{m}'_t\})$ is the conditional Shannon entropy. A similar bound was derived in Ref. [10] for the case without measurement errors.

In the next step, we relate the Shannon entropy to the coherent information. To this end, we rewrite the density matrices $\rho_{Q\mathbf{M}}$ and $\rho_{Q\mathbf{M}R}$ in the error configuration expansion in terms of the syndromes $\boldsymbol{m}_T$ and $\boldsymbol{m}'_t$

$$
\begin{aligned}
\rho_{Q\mathbf{M}} &= \sum_{\boldsymbol{m}_T, \{\boldsymbol{m}'_t\}} \frac{1}{2^K} \Pr(\boldsymbol{m}_T, \{\boldsymbol{m}'_t\}) \mathcal{P}_{\boldsymbol{m}_T} \otimes_t |\boldsymbol{m}'_t\rangle\langle\boldsymbol{m}'_t|, \\
\rho_{Q\mathbf{M}R} &= \sum_{\boldsymbol{\kappa}, \boldsymbol{m}_T, \{\boldsymbol{m}'_t\}} \Pr(\boldsymbol{\kappa}, \boldsymbol{m}_T, \{\boldsymbol{m}'_t\}) \\
&\qquad \mathcal{O}_0 \overline{g}^{\boldsymbol{\kappa}} |\Phi\rangle\langle\Phi|_{RQ} \overline{g}^{\boldsymbol{\kappa}} \mathcal{O}_0 \otimes_t |\boldsymbol{m}'_t\rangle\langle\boldsymbol{m}'_t|,
\end{aligned}
\tag{25}
$$

where $\Pr(\boldsymbol{m}_T, \{\boldsymbol{m}'_t\})$ $(\Pr(\boldsymbol{\kappa}, \boldsymbol{m}_T, \{\boldsymbol{m}'_t\}))$ is the total probability of error operators that are consistent with the syndrome (and are in the class $\overline{g}^{\boldsymbol{\kappa}}$). Here, we decompose the cumulative error as $\mathcal{O}_{\Sigma \boldsymbol{b}_T} := \mathcal{O}_0 \overline{g}^{\boldsymbol{c}}$, where $\mathcal{O}_0$ is a reference error operator that creates the syndrome $\boldsymbol{m}_T$, and $|\Phi\rangle_{RQ}$ is the maximally entangled state between the reference R and the system Q, which contains K Bell pairs. It follows that

$$
S(\rho_{Q\mathbf{M}}) = H(\boldsymbol{m}_T, \{\boldsymbol{m}'_t\}) + K \log 2, \tag{26}
$$
$$
S(\rho_{Q\mathbf{M}R}) = H(\boldsymbol{\kappa}, \boldsymbol{m}_T, \{\boldsymbol{m}'_t\}). \tag{27}
$$

Thus, we have the relation

$$
\overline{\Delta} \leq H(\boldsymbol{\kappa}|\boldsymbol{m}_T, \{\boldsymbol{m}'_t\}) = K \log 2 - I_c(R\rangle Q\mathbf{M}). \tag{28}
$$

In conclusion, we have shown that the ML decoder is asymptotically optimal, i.e. it has a vanishing failure probability as long as the coherent information approaches its maximum in the thermodynamic limit. We remark that the proof works for decoherence described by Pauli channels, and it does not apply to cases with coherent errors.

In Sec. IV, we show that the stat-mech model derived for coherent information is dual to that obtained for the optimal decoding algorithm, which is consistent with the rigorous relation shown in this section.

IV. STATISTICAL MECHANICS MODEL

In this section, we develop stat-mech descriptions for the Rényi versions of information diagnostics in our quantum memory dynamics. The key step is to map $\operatorname{tr} \rho_{Q\mathbf{M}}^n$ to the partition function of a stat-mech model. We then find that the n-th moments of different density matrices map to the same model but with different boundary conditions or defect insertions. Furthermore, the Rényi versions of the information diagnostics are identified with distinct observables in the stat-mech model.

In Sec. IV A, we introduce the stabilizer expansion of $\rho_{Q\mathbf{M}}$ and use it to obtain a stat-mech model consisting of $n - 1$ flavors of Ising spins. The model exhibits symmetries determined by the redundancies $\mathcal{R}$ in the check operators. For example, in the surface code the check operators exhibit a global $\mathbb{Z}_2$ redundancy and the decoding transition is associated with a $\mathbb{Z}_2$ symmetry-breaking transition in the stat-mech model. In this case, the measurements of check operators suppress the ferromagnetic ordering and therefore increase the decoherence threshold.

Our stat-mech model is dual to the model obtained for the maximum likelihood decoder [3, 19, 36]. In Sec. IV B, we consider an alternative expansion of the same density matrix in terms of error configurations. We then obtain a stat-mech model with the same partition function as (i.e. dual to) the model in the stabilizer expansion. We show that our stat-mech model in the error configuration expansion reduces to the one associated with the decoding fidelity of the ML decoder in the limit $n \to 1$.

We remark that the mapping to stat-mech models based on the two expansions of the density matrix and the duality between the resulting models are first proposed in Ref. [4] to study the information diagnostics in topological quantum memory subject to local decoherence. Our stabilizer expansion generalizes the "loop expansion" in Ref. [4] and allows better identification of the symmetry in the resulting stat-mech model.

A. Stat-mech models in the stabilizer expansion

Developing stat-mech models in the stabilizer expansion takes two steps. In Sec. IV A 1, we express the density matrix at time T as a sum of operators weighted by real positive numbers. Then, in Sec. IV A 2, we map the information diagnostics, the coherent information and relative entropy, to the partition functions of stat-mech models.

1. State evolution under noise and imperfect measurements

Here, we formulate the evolution of the density matrix $\rho_{Q\mathbf{M}}$ under T layers of decoherence and noisy syndrome measurements in the stabilizer expansion. We consider the initial state $\rho_{Q\mathbf{M}_0}$ to be the maximally mixed state in the code space of a stabilizer code $\mathcal{S}$

$$
\rho_{Q\mathbf{M}_0} = \frac{1}{2^K} \prod_{i \in I} \frac{1 + g_i}{2} = \frac{1}{2^N} \sum_{\boldsymbol{u}} g^{\boldsymbol{u}}, \tag{29}
$$

where $\boldsymbol{u} \in \mathbb{Z}_2^I$ is an I-component binary vector that labels the stabilizer $g^{\boldsymbol{u}} \in \mathcal{S}$ (as in Eq. (12)). From now on, we do not keep track of the overall prefactor in the expansion, which can always be restored by enforcing $\operatorname{tr} \rho = 1$.

The evolution of $\rho_{Q\mathbf{M}_0}$ under physical decoherence takes a simple form in the stabilizer expansion because Pauli operators acquire only overall prefactors under Pauli channels, i.e.

$$
\mathcal{N}[\rho_{Q\mathbf{M}_0}] = \sum_{\boldsymbol{u}} e^{-H(\boldsymbol{u})} g^{\boldsymbol{u}}. \tag{30}
$$

To explicitly write down the prefactor, we express the stabilizer as

$$g_i = i^{\boldsymbol{a}_i^x \cdot \boldsymbol{a}_i^z} X^{\boldsymbol{a}_i^x} Z^{\boldsymbol{a}_i^z}, \qquad (31)$$

where $\boldsymbol{a}_i^{x(z)}$ is a binary vector of length N specifying the support of Pauli-X(Z) on physical qubits.

The prefactor then takes the form

$$
\begin{aligned}
&e^{-H(\boldsymbol{u})} \\
&= \prod_r \left[\left(1 - \sum_\alpha p_\alpha \right) + p_x h_r^z + p_z h_r^x + p_y h_r^x h_r^z \right], \quad (32)
\end{aligned}
$$

where the binary variable $h_r^{x(z)} := \prod_i (-1)^{u_i a_{i,r}^{x(z)}}$ tracks the support of Pauli-X(Z) at site r. This prefactor can be viewed as the Boltzmann weight of a classical Hamiltonian

$$H(\boldsymbol{u}) = \sum_r -J_z h_r^x - J_x h_r^z - J_y h_r^x h_r^z \qquad (33)$$

up to a constant term independent of $\boldsymbol{u}$, with

$$J_{\alpha_1} = -\frac{1}{4} \log \frac{(1 - 2p_{\alpha_1} - 2p_{\alpha_2})(1 - 2p_{\alpha_1} - 2p_{\alpha_3})}{1 - 2p_{\alpha_2} - 2p_{\alpha_3}}, \quad (34)$$

where $\alpha_{1,2,3}$ are different labels chosen from $\{x, y, z\}$.

Next, we obtain the evolution of ρ_{QM} under noisy syndrome measurements. Each Pauli operator $g^{\boldsymbol{u}}$ in the stabilizer expansion is evolved under the measurement channel $\mathcal{M}_{t,i}$ associated with measuring g_i at time t, which takes the form

$$
\begin{aligned}
\mathcal{M}_{t,i}[g^{\boldsymbol{u}}] &= \sum_{m_{t,i}=0,1} \mathcal{P}_{m_{t,i}} g^{\boldsymbol{u}} \mathcal{P}_{m_{t,i}} \mathcal{N}_M[|m_{t,i}\rangle\langle m_{t,i}|] \\
&= \frac{1}{2} g^{\boldsymbol{u}} \left(1 + e^{-\mu_q} g_i Z_{t,i} \right), \quad (35)
\end{aligned}
$$

where $\mathcal{P}_{m_{t,i}} = (1 + (-1)^{m_{t,i}} g_i)/2$ is the projector onto corresponding eigenstates, $\mu_q = -\log(1 - 2q)$, and $Z_{t,i}$ is the Pauli-Z operator for the measurement ancilla. After a full layer of stabilizer measurements at the first time step, the resulting density matrix is

$$\rho_{QM_1} = \sum_{\boldsymbol{u}_0, \boldsymbol{u}_1} e^{-H(\boldsymbol{u}_0) - \mu_q |\boldsymbol{u}_1 - \boldsymbol{u}_0|} g^{\boldsymbol{u}_1} Z_1^{\boldsymbol{u}_1 - \boldsymbol{u}_0}, \qquad (36)$$

where $|\boldsymbol{u}| = \sum_i u_i$ counts the number of nonzero elements in the binary vector $\boldsymbol{u}$, and $Z_1^{\boldsymbol{u}} := \otimes_{i=1}^I Z_{1,i}^{u_i}$ is a product of Pauli-Z on ancillae. We remark that $|\boldsymbol{u}_1 - \boldsymbol{u}_0|$ serves as an Ising coupling between spins at consecutive times.

After T layers of decoherence and noisy measurements, the density matrix ρ_{QM} takes the form

$$\rho_{QM} = \sum_{\{\boldsymbol{u}_t\}} e^{-\mathcal{H}(\{\boldsymbol{u}_t\})} g^{\boldsymbol{u}_T} \otimes_{t=1}^T Z_t^{\boldsymbol{u}_t - \boldsymbol{u}_{t-1}}, \qquad (37)$$

where the weight of the Pauli operator can be written as the Boltzmann weight associated with a classical Hamiltonian

$$\mathcal{H}(\{\boldsymbol{u}_t\}) = \sum_{t=1}^T H(\boldsymbol{u}_{t-1}) + \mu_q |\boldsymbol{u}_t - \boldsymbol{u}_{t-1}|. \qquad (38)$$

Here, the Hamiltonian involves $T + 1$ layers of classical variables with intralayer coupling described by $H(\boldsymbol{u}_t)$ in the first T layers and interlayer coupling $\mu_q |\boldsymbol{u}_t - \boldsymbol{u}_{t-1}|$ controlled by the measurement error rate. We note that in the case that $\mathcal{S}$ has redundancies, a different set of binary vectors $\{\boldsymbol{u}_t'\}$ related to $\{\boldsymbol{u}_t\}$ via a linear transformation can represent the same Pauli operator $g^{\boldsymbol{u}_t}$. Here, all $\boldsymbol{u}_t$ must transform together to keep the Pauli-Z operator on the ancillae, i.e. $|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}|$, invariant. The transformation, therefore, relates two sets of binary vectors with the same weight and is a symmetry of the Hamiltonian. This indicates that the Hamiltonian $\mathcal{H}(\{\boldsymbol{u}_t\})$ exhibits a symmetry group $\mathcal{R}$ given by the redundancy of the stabilizers. In addition, we note that the Hamiltonian can exhibit "spacetime" symmetries arising from the symmetries of the code and decoherence model under translations, rotations, etc.

Examples.— We here discuss the Hamiltonian $\mathcal{H}(\boldsymbol{u})$ in a few examples. In the 2D toric code (the surface code on the 2D square lattice with periodic boundary conditions), the qubits are defined on the edges, and the model involves classical variables associated with check operators on the vertices and plaquettes. The Pauli-X(Z) operator for each qubit is only involved in two check operators. Accordingly, $h_r^{x(z)}$ in the Hamiltonian $H(\boldsymbol{u})$ is a two-body term, and $h_r^x h_r^z$ is a four-body term. The check operators in the toric code have redundancies as the product of all plaquette (star) check operators is the identity, indicating an $\mathcal{R} = \mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry of the Hamiltonian.

In contrast, the 2D planar surface code with open boundary conditions does not have extraneous check operators. This indicates that the Hamiltonian does not possess an exact $\mathcal{R}$ symmetry. Nevertheless, the product of all plaquette (star) check operators in the planar code cancels in the bulk with non-trivial support only on the boundary, indicating an approximate symmetry that is broken only on the boundary. We note that such approximate symmetry is still crucial for governing the potential symmetry-breaking transition in the thermodynamic limit.

In the 2D repetition code on a square lattice with periodic boundary conditions, each qubit (defined on vertices) is involved in four ZZ-check operators, leading to a four-body coupling h_r^x in the Hamiltonian. Due to the local redundancies in the check operators, namely that the product of four checks around a plaquette is the identity, the Hamiltonian exhibits an extensive number of $\mathbb{Z}_2$ symmetries, whose local structure allows one to identify $H(\boldsymbol{u})$ as a $\mathbb{Z}_2$ gauge theory.

2. Stat-mech models for information diagnostics

We are now ready to develop stat-mech models for coherent information, decoding fidelity, and relative entropy. The key step is to identify the n-th moment of the density matrix, e.g. $\operatorname{tr} \rho_{Q\mathbf{M}}^n$, with the partition function of a $(n-1)$-flavor Ising model,

$$\mathcal{Z}_n := \operatorname{tr} \rho_{Q\mathbf{M}}^n. \tag{39}$$

In the stabilizer expansion, the n-th moment $\operatorname{tr} \rho_{Q\mathbf{M}}^n = \mathcal{Z}_n$ involves a factor

$$
\Omega(\{\boldsymbol{u}^{\mathsf{a}}\}) := \operatorname{tr}\left(\prod_{\mathsf{a}=1}^n g^{\boldsymbol{u}_T^{\mathsf{a}}} \prod_{t=1}^T Z_t^{\boldsymbol{u}_t^{\mathsf{a}} - \boldsymbol{u}_{t-1}^{\mathsf{a}}}\right)
$$
$$
= \sum_{\boldsymbol{r} \in \mathcal{R}} \prod_{t=0}^T \delta\left(\sum_{\mathsf{a}=1}^n \boldsymbol{u}_t^{\mathsf{a}} = \boldsymbol{r}\right), \tag{40}
$$

where $\boldsymbol{u}^{\mathsf{a}}$ is the binary variable associated with the a-th copy of the density matrix $\rho_{Q\mathbf{M}}$. Here, Ω is nonvanishing only if $\sum_{\mathsf{a}} \boldsymbol{u}_t^{\mathsf{a}} \in \mathcal{R}$ for all $0 \le t \le T$. We note that we have omitted an overall constant in the second equality.

The delta function in Eq. (40) allows one to write the n-th moment as a partition function of classical variables in the first $n-1$ copies

$$
\mathcal{Z}_n = \sum_{\boldsymbol{r} \in \mathcal{R}} \sum_{\{\boldsymbol{u}_t^{\mathsf{a}}\}} e^{-\sum_{\mathsf{a}=1}^{n-1} \mathcal{H}(\{\boldsymbol{u}_t^{\mathsf{a}}\}) - \mathcal{H}(\{\sum_{\mathsf{a}=1}^{n-1} \boldsymbol{u}_t^{\mathsf{a}} + \boldsymbol{r}\})}. \tag{41}
$$

We note that since $\mathcal{H}(\{\boldsymbol{u}_t + \boldsymbol{r}\}) = \mathcal{H}(\{\boldsymbol{u}_t\})$ for $\boldsymbol{r} \in \mathcal{R}$, the summation over the redundancy group gives only an overall prefactor and therefore can be omitted.

It is convenient to introduce classical Ising spin variables $\sigma_{t,i}^{\mathsf{a}} := (-1)^{u_{t,i}^{\mathsf{a}}}$ to re-express the partition function as

$$
\mathcal{Z}_n = \sum_{\boldsymbol{\sigma}^1, \dots, \boldsymbol{\sigma}^{n-1}} e^{-\mathcal{H}^{(n)}(\{\boldsymbol{\sigma}^{\mathsf{a}}\})} \tag{42}
$$

with $\mathcal{H}^{(n)}$ being the Hamiltonian of $(n-1)$-flavor Ising spins,

$$
\mathcal{H}^{(n)}(\{\boldsymbol{\sigma}^{\mathsf{a}}\}) := \sum_{\mathsf{a}=1}^{n-1} \mathcal{H}(\boldsymbol{\sigma}^{\mathsf{a}}) + \mathcal{H}\left(\prod_{\mathsf{a}=1}^{n-1} \boldsymbol{\sigma}^{\mathsf{a}}\right). \tag{43}
$$

We note that the Hamiltonian $\mathcal{H}^{(n)}$ has open boundary conditions on the final $t = T$ layer.

The symmetry of $\mathcal{H}^{(n)}$ is convenient to see by formulating the Hamiltonian $\mathcal{H}^{(n)}$ as n copies of single-flavor Hamiltonian $\mathcal{H}(\boldsymbol{\sigma}^{\mathsf{a}})$ subject to the constraint $\boldsymbol{\sigma}^n = \prod_{\mathsf{a}=1}^{n-1} \boldsymbol{\sigma}^{\mathsf{a}}$. Then, the symmetry of $\mathcal{H}^{(n)}$ contains n copies of the symmetry $\mathcal{R}$ of the single-flavor Hamiltonian. The $\mathcal{R}^n$ symmetry is further extended by the S_n symmetry permuting over n identical copies of $\mathcal{H}(\boldsymbol{\sigma}^{\mathsf{a}})$. After factoring out a copy of symmetry $\mathcal{R}$ due to the constraint, we obtain the global symmetry $(\mathcal{R}^n \rtimes S_n)/\mathcal{R}$ of the Hamiltonian $\mathcal{H}^{(n)}$. Again, we note that when evaluating the

thermodynamic properties in the stat-mech model, one also needs to involve the approximate symmetries and the spacetime symmetries as a part of $\mathcal{R}$.

We remark that the stat-mech model reduces to a single-flavor model for $n = 2$ and $n \to \infty$. At $n = 2$, the Hamiltonian $\mathcal{H}^{(2)} = 2\mathcal{H}(\boldsymbol{\sigma}^1)$ contains a single flavor. At $n \to \infty$, the interaction $\mathcal{H}(\prod_{\mathsf{a}} \boldsymbol{\sigma}^{\mathsf{a}})$ between flavors becomes negligible[4], and the system consists of $n-1$ independent copies of $\mathcal{H}(\boldsymbol{\sigma}^{\mathsf{a}})$. These two cases provide useful intuition regarding the potential phases and phase transitions in the $n \to 1$ limit, despite different universalities due to different symmetries of the model.

Coherent information.— To calculate the Rényi coherent information, we now map the n-th moment $\operatorname{tr} \rho_{Q\mathbf{M}R}^n$ to the stat-mech model and show that the coherent information corresponds to the excess free energy of inserting defects in the stat-mech model.

To begin, the initial state $\rho_{Q\mathbf{M}_0 R}$ is the maximally entangled state between the code space of the system Q and the reference R that contains K qubits

$$
\rho_{Q\mathbf{M}_0 R} = \prod_{k=1}^K \frac{1 + s_k^x \bar{g}_k^x}{2} \frac{1 + s_k^z \bar{g}_k^z}{2} \prod_{i=1}^I \frac{1 + g_i}{2}
$$
$$
= \frac{1}{2^{K+N}} \sum_{\boldsymbol{u}, \boldsymbol{\kappa}} g^{\boldsymbol{u}} \bar{g}^{\boldsymbol{\kappa}} s^{\boldsymbol{\kappa}}, \tag{44}
$$

where $s_k^{x(z)}$ is the Pauli-X(Z) operator for the k-th qubit in the reference R. The maximally entangled state is stabilized by $s_k^x \bar{g}_k^x$ and $s_k^z \bar{g}_k^z$. Each term in the stabilizer expansion can be labeled by the binary vector $\boldsymbol{u}$ together with a $2K$-component binary vector $\boldsymbol{\kappa}$ which labels the content in the logical space and reference. We will omit the constant prefactor in the second equality in what follows.

Under the time evolution of quantum memory, the density matrix $\rho_{Q\mathbf{M}R}$ evolves as

$$
\rho_{Q\mathbf{M}R} = \sum_{\{\boldsymbol{u}_t\}, \boldsymbol{\kappa}} e^{-\mathcal{H}_{\boldsymbol{\kappa}}(\{\boldsymbol{u}_t\})} g^{\boldsymbol{u}_T} \bar{g}^{\boldsymbol{\kappa}} s^{\boldsymbol{\kappa}} \otimes_{t=1}^T Z_t^{\boldsymbol{u}_t - \boldsymbol{u}_{t-1}}. \tag{45}
$$

Here, the logical operator $\bar{g}^{\boldsymbol{\kappa}}$ and the Pauli operator $s^{\boldsymbol{\kappa}}$ on the reference do not evolve, and the terms in the expansion acquire a weight $e^{-\mathcal{H}_{\boldsymbol{\kappa}}}$ that depends on the logical operator $\boldsymbol{\kappa}$.

The Hamiltonian that governs the weight is given by

$$
\mathcal{H}_{\boldsymbol{\kappa}}(\{\boldsymbol{u}_t\}) = \sum_{t=1}^T H_{\boldsymbol{\kappa}}(\boldsymbol{u}_{t-1}) + \mu_q |\boldsymbol{u}_t - \boldsymbol{u}_{t-1}|, \tag{46}
$$

$$
H_{\boldsymbol{\kappa}}(\boldsymbol{u}) = \sum_r -J_z h_{\boldsymbol{\kappa},r}^x - J_x h_{\boldsymbol{\kappa},r}^z - J_y h_{\boldsymbol{\kappa},r}^x h_{\boldsymbol{\kappa},r}^z, \tag{47}
$$

---

[4] One can view the $n-1$ flavors of the Ising spin as an additional dimension. Let $\boldsymbol{\tau}^{\mathsf{a}} := \prod_{\mathsf{b}=\mathsf{a}}^{n-1} \boldsymbol{\sigma}^{\mathsf{b}}$. The Hamiltonian is $\mathcal{H}^{(n)} = \mathcal{H}(\boldsymbol{\tau}^1) + \sum_{\mathsf{a}=1}^{n-2} \mathcal{H}(\boldsymbol{\tau}^{\mathsf{a}} \boldsymbol{\tau}^{\mathsf{a}+1})$. The inter-replica coupling becomes $\mathcal{H}(\boldsymbol{\tau}^1)$, a boundary field in the flavor dimension, which can be neglected in the limit $n \to \infty$.

where

$$h^x_{\boldsymbol{\kappa},r}(\boldsymbol{u}) := (-1)^{\sum_{k=1}^{2K} \kappa_k \overline{a}^x_{k,r}} h^x_r(\boldsymbol{u}), \qquad (48)$$

$$h^z_{\boldsymbol{\kappa},r}(\boldsymbol{u}) := (-1)^{\sum_{k=1}^{2K} \kappa_k \overline{a}^z_{k,r}} h^z_r(\boldsymbol{u}). \qquad (49)$$

Here, $\overline{\boldsymbol{a}}_k$ is the $2N$-component binary vector associated with the $2K$ logical (X and Z) operators. The logical operator $\overline{g}^{\boldsymbol{\kappa}}$ in the expansion therefore controls the sign of $h^{x,z}_r$ effectively inserting a defect in the stat-mech model along the temporal direction.

Next, we consider the n-th moment $\operatorname{tr}\rho^n_{QMR}$ in the stabilizer expansion. The trace is non-vanishing only if $\boldsymbol{\kappa}^n = \sum_{\mathsf{a}=1}^{n-1} \boldsymbol{\kappa}^{\mathsf{a}}$. The n-th moment thus takes the form

$$\operatorname{tr}\rho^n_{QMR} = \frac{1}{2^{(n-1)K}} \sum_{\{\boldsymbol{\kappa}^{\mathsf{a}}\}} \mathcal{Z}_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\}), \qquad (50)$$

which is a summation of the partition functions with different defect insertions labeled by $\boldsymbol{\kappa}^{\mathsf{a}}$:

$$\mathcal{Z}_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\}) = \sum_{\{\boldsymbol{u}^{\mathsf{a}}\}} e^{-\sum_{\mathsf{a}=1}^{n-1} \mathcal{H}_{\boldsymbol{\kappa}^{\mathsf{a}}}(\boldsymbol{u}^{\mathsf{a}}) - \mathcal{H}_{\sum_{\mathsf{a}} \boldsymbol{\kappa}^{\mathsf{a}}}(\sum_{\mathsf{a}} \boldsymbol{u}^{\mathsf{a}})}. \qquad (51)$$

We can thus write down the Rényi-n coherent information as

$$I^{(n)}_c(R\rangle Q\mathbf{M}) = \frac{1}{n-1} \log\left(\sum_{\{\boldsymbol{\kappa}^{\mathsf{a}}\}} e^{-\Delta F(\{\boldsymbol{\kappa}^{\mathsf{a}}\})}\right) - K\log 2, \qquad (52)$$

where $\Delta F(\{\boldsymbol{\kappa}^{\mathsf{a}}\}) := -\log \mathcal{Z}_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\}) + \log \mathcal{Z}_n$ is the excess free energy of inserting the defect.

Relative entropy.— To evaluate the relative entropy, we first express $\rho_{Q\mathbf{M}_0,\boldsymbol{s}} = w_{\boldsymbol{s}} \rho_{Q\mathbf{M}_0} w^\dagger_{\boldsymbol{s}}$ in the stabilizer expansion

$$\rho_{Q\mathbf{M}_0,\boldsymbol{s}} = \sum_{\boldsymbol{u}} (-1)^{\boldsymbol{s}\cdot\boldsymbol{u}} g^{\boldsymbol{u}}, \qquad (53)$$

where we recall that $\boldsymbol{s}$ is a nontrivial syndrome, resulting from Pauli operator $w_{\boldsymbol{s}}$.

After evolving for T steps, the density matrix $\rho_{Q\mathbf{M},\boldsymbol{s}}$ takes the form

$$\rho_{Q\mathbf{M},\boldsymbol{s}} = \sum_{\{\boldsymbol{u}_t\}} (-1)^{\boldsymbol{s}\cdot\boldsymbol{u}_0} e^{-\mathcal{H}(\{\boldsymbol{u}_t\})} g^{\boldsymbol{u}_T} \otimes_{t=1}^{T} Z^{\boldsymbol{u}_t - \boldsymbol{u}_{t-1}}_t. \qquad (54)$$

Accordingly, the n-th moment $\operatorname{tr}\rho_{Q\mathbf{M}} \rho^{n-1}_{Q\mathbf{M},\boldsymbol{s}}$ is given by

$$\operatorname{tr}\rho_{Q\mathbf{M}} \rho^{n-1}_{Q\mathbf{M},\boldsymbol{s}} = \sum_{\{\boldsymbol{u}^{\mathsf{a}}\}} (-1)^{\boldsymbol{s}\cdot\boldsymbol{u}^1_0} e^{-\mathcal{H}^{(n)}(\{\boldsymbol{u}^{\mathsf{a}}\})}. \qquad (55)$$

We note that the phase factor is inserted for the last $n-1$ copies, and we have used the constraint $\boldsymbol{u}^1_0 = \sum_{\mathsf{a}=2}^{n} \boldsymbol{u}^{\mathsf{a}}_0$ to simplify the expression.

We can thus write down the relative entropy as a boundary correlation function between classical spins in the first copy

$$D^{(n)}_{\boldsymbol{s}} = -\frac{1}{n-1} \log\left\langle \prod_i (\sigma^1_{0,i})^{s_i} \right\rangle, \qquad (56)$$

where $\langle\cdot\rangle$ denotes the expectation value in the defect-free stat-mech model of $n-1$ flavors, and we recall that $\sigma^{\mathsf{a}}_{t,i} = (-1)^{u^{\mathsf{a}}_{t,i}}$.

KL divergence.— Last, we consider the Rényi-n version of the KL divergence in the stat-mech model. The key difference is that we trace over the system Q to obtain the reduced density matrix of the measurement ancillae $\rho_{\mathbf{M}} = \operatorname{tr}_Q \rho_{Q\mathbf{M}}$. The trace is non-vanishing only if $g^{\boldsymbol{u}_T}$ is the identity, i.e $\boldsymbol{u}_T \in \mathcal{R}$. We note that for any $\boldsymbol{u}_T \in \mathcal{R}$, one can shift $\boldsymbol{u}_t \mapsto \boldsymbol{u}_t - \boldsymbol{u}_T$ while keeping the weight invariant in the stabilizer expansion. Thus, without loss of generality, we can set $\boldsymbol{u}_T = 0$. Then, the expansion takes the form

$$\rho_{\mathbf{M}} = \sum_{\boldsymbol{u}_t} e^{-\mathcal{H}^+(\{\boldsymbol{u}_t\})} Z^{-\boldsymbol{u}_{T-1}}_T \otimes_{t=1}^{T-2} Z^{\boldsymbol{u}_t - \boldsymbol{u}_{t-1}}_t, \qquad (57)$$

where

$$\mathcal{H}^+(\{\boldsymbol{u}_t\}) = \sum_{t=1}^{T} H(\boldsymbol{u}_{t-1}) + \sum_{t=1}^{T-1} \mu_q|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}| + \mu_q|\boldsymbol{u}_{T-1}|. \qquad (58)$$

Here, the last term changes under the symmetry transformation $\boldsymbol{u}_t \mapsto \boldsymbol{u}_t + \boldsymbol{v}$ with $\boldsymbol{v} \in \mathcal{R}$, and therefore is a symmetry-breaking field at the top boundary $t = T$ in the stat-mech model. Accordingly, the n-th moment $\operatorname{tr}\rho^n_{\mathbf{M}}$ takes the form

$$\operatorname{tr}\rho^n_{\mathbf{M}} = \sum_{\{\boldsymbol{u}^{\mathsf{a}}\}} e^{-\mathcal{H}^{(n),+}(\{\boldsymbol{u}^{\mathsf{a}}\})}$$
$$= \sum_{\{\boldsymbol{u}^{\mathsf{a}}\}} e^{\sum_{\mathsf{a}=1}^{n-1} -\mathcal{H}^+(\boldsymbol{u}^{\mathsf{a}}) - \mathcal{H}^+(\sum_{\mathsf{a}=1}^{n-1} \boldsymbol{u}^{\mathsf{a}})}. \qquad (59)$$

Similar to the quantum relative entropy, the Rényi-n KL divergence is given by

$$D^{(n)}_{\mathrm{KL},\boldsymbol{s}} = -\frac{1}{n-1} \log\left\langle \prod_i (\sigma^1_{0,i})^{s_i} \right\rangle_+, \qquad (60)$$

where $\langle\cdot\rangle_+$ denotes expectation value in the defect-free stat-mech model of $n-1$ flavors with symmetry-breaking boundary conditions.

The existence of symmetry-breaking boundary fields may lead to qualitative distinctions between D_{KL} and D, linked to the subtle difference between successful error correction and retaining quantum information in $\rho_{Q\mathbf{M}}$. In the example of the 2D surface code, the boundary field always orders the spins in the multi-flavor Ising model (derived in Appendix D) when $t = O(1)$, leading

to $D_{\text{KL}} = O(1)$. This indicates that noisy measurements cannot identify distinct error syndromes after $O(1)$ time, although the coherent quantum information $I_c(R\rangle Q\mathbf{M})$ may remain maximal for sufficiently small error rates. This highlights the need for $t = O(L)$ rounds of noisy measurements for successful error correction. We explain the same physics in the 3D RPIM for the ML decoder in Appendix F.

B. Stat-mech model in the error configuration expansion

In this section, we take the error configuration expansion of the mixed density matrix $\rho_{Q\mathbf{M}}$ in Eq. (21) and map the n-th moment $\operatorname{tr}\rho_{Q\mathbf{M}}^n$ to the partition function of a stat-mech model, which is dual to the model obtained in the stabilizer expansion (Eq. (42)). In the limit of $n \to 1$, the stat-mech model in the error configuration expansion reduces to a disordered model on the Nishimori line, which governs the decoding fidelity of the maximum likelihood decoder.

First, we use the error configuration expansion in Eq. (21) to express the n-th moment $\operatorname{tr}\rho_{Q\mathbf{M}}^n$ as

$$\operatorname{tr}\rho_{Q\mathbf{M}}^n = \sum_{\{\boldsymbol{b}_t^{\mathsf{a}}, \boldsymbol{\epsilon}_t^{\mathsf{a}}\}} \prod_{\mathsf{a}=1}^n \Pr(\{\boldsymbol{b}_t^{\mathsf{a}}, \boldsymbol{\epsilon}_t^{\mathsf{a}}\}) \prod_{t,\mathsf{a}} \delta_{\boldsymbol{m}_t^{\mathsf{a}}+\boldsymbol{\epsilon}_t^{\mathsf{a}}, \boldsymbol{m}_t^{\mathsf{a}+1}+\boldsymbol{\epsilon}_t^{\mathsf{a}+1}}$$
$$\operatorname{tr}\left[\prod_{\mathsf{a}=1}^n \mathcal{O}_{\Sigma\boldsymbol{b}_T^{\mathsf{a}}}\, \rho_0\, \mathcal{O}_{\Sigma\boldsymbol{b}_T^{\mathsf{a}}}^\dagger\right], \qquad (61)$$

where $\Sigma\boldsymbol{b}_T^{\mathsf{a}} := \sum_{\tau=1}^T \boldsymbol{b}_\tau^{\mathsf{a}}$ is a shorthand notation, and we again choose ρ_0 to be the maximally mixed state in the logical space. In this expansion, the n-th moment becomes a constrained sum over error configurations weighted by their probability.

The error configuration has a non-vanishing weight only if the following two conditions are satisfied: (1) the physical and measurement errors in each copy produce the same error syndromes at every time step t, i.e.

$$\boldsymbol{m}_t^1 + \boldsymbol{\epsilon}_t^1 = \boldsymbol{m}_t^{\mathsf{a}} + \boldsymbol{\epsilon}_t^{\mathsf{a}}, \text{ for } \mathsf{a} = 2, 3, \dots, n; \qquad (62)$$

(2) the cumulative physical errors $\Sigma\boldsymbol{b}_T^{\mathsf{a}}$ produce the same syndromes in each copy, i.e.

$$\boldsymbol{m}_T^1 = \boldsymbol{m}_T^{\mathsf{a}}, \text{ for } \mathsf{a} = 2, 3, \dots, n. \qquad (63)$$

Here, $\boldsymbol{m}_t^{\mathsf{a}}$ is the true eigenvalue of measured stabilizers and is flipped only by the physical error operators,

$$m_{t,i}^{\mathsf{a}} - m_{t-1,i}^{\mathsf{a}} = \langle\langle \boldsymbol{a}_i, \boldsymbol{b}_t^{\mathsf{a}}\rangle\rangle, \qquad (64)$$

where $\langle\langle \boldsymbol{a}, \boldsymbol{b}\rangle\rangle = \boldsymbol{a}^x \cdot \boldsymbol{b}^z + \boldsymbol{a}^z \cdot \boldsymbol{b}^x$ is the binary symplectic form. Hence, the first condition requires

$$\langle\langle \boldsymbol{a}_i, \boldsymbol{b}_t^1\rangle\rangle + \epsilon_{t,i}^1 - \epsilon_{t-1,i}^1 = \langle\langle \boldsymbol{a}_i, \boldsymbol{b}_t^{\mathsf{a}}\rangle\rangle + \epsilon_{t,i}^{\mathsf{a}} - \epsilon_{t-1,i}^{\mathsf{a}}. \qquad (65)$$

We note that $\boldsymbol{m}_0^{\mathsf{a}} = \boldsymbol{\epsilon}_0^{\mathsf{a}} = \boldsymbol{0}$ in Eqs. (64) and (65) for $t = 1$. The second condition requires the same readout

errors in the last round of measurements in each copy, i.e. $\boldsymbol{\epsilon}_T^1 = \boldsymbol{\epsilon}_T^{\mathsf{a}}$.

Before proceeding, we provide an intuitive understanding of Eq. (65). One can view the symplectic form $\langle\langle \boldsymbol{a}_i, \boldsymbol{b}_t\rangle\rangle = \partial\boldsymbol{b}_t$ as the boundary of the physical error chain $\boldsymbol{b}_t$. The difference in measurement error $\epsilon_{t,i} - \epsilon_{t-1,i}$ is the boundary $\partial\boldsymbol{\epsilon}_t$ of measurement error chain $\boldsymbol{\epsilon}_t$. Equation (65) then takes the form

$$\partial(\boldsymbol{b}_t^1 - \boldsymbol{b}_t^{\mathsf{a}}) + \partial(\boldsymbol{\epsilon}_t^1 - \boldsymbol{\epsilon}_t^{\mathsf{a}}) = 0. \qquad (66)$$

This indicates the (physical and measurement) error configuration in the first and the a-th copy form a closed chain.

We can thus introduce the new variables $\boldsymbol{e}_t^{\mathsf{a}}$ to specify the closed chains and label the configuration of $(\boldsymbol{b}_t^{\mathsf{a}}, \boldsymbol{\epsilon}_t^{\mathsf{a}})$ with non-vanishing weights:

$$\boldsymbol{b}_t^{\mathsf{a}} - \boldsymbol{b}_t^1 = \boldsymbol{e}_t^{\mathsf{a}}, \qquad (67)$$
$$\delta\epsilon_{t,i}^{\mathsf{a}} - \delta\epsilon_{t,i}^1 = \langle\langle \boldsymbol{a}_i, \boldsymbol{e}_t^{\mathsf{a}}\rangle\rangle. \qquad (68)$$

Because each copy has the same syndrome $\boldsymbol{m}_T^1 = \boldsymbol{m}_T^{\mathsf{a}}$ at the final step, the cumulative physical errors in different copies can only differ by a stabilizer $\boldsymbol{a}_{\boldsymbol{v}}^{\mathsf{a}} := \sum_{i=1}^{I_0} v_i^{\mathsf{a}} \boldsymbol{a}_i$ and a logical operator $\overline{\boldsymbol{a}}_{\boldsymbol{\kappa}}^{\mathsf{a}} := \sum_{k=1}^{2K} \kappa_k^{\mathsf{a}} \overline{\boldsymbol{a}}_k$

$$\Sigma\boldsymbol{e}_T^{\mathsf{a}} = \Sigma\boldsymbol{b}_T^{\mathsf{a}} - \Sigma\boldsymbol{b}_T^1 = \boldsymbol{a}_{\boldsymbol{v}}^{\mathsf{a}} + \overline{\boldsymbol{a}}_{\boldsymbol{\kappa}}^{\mathsf{a}}. \qquad (69)$$

Thus, one can divide the error chain $(\boldsymbol{b}_t^{\mathsf{a}}, \boldsymbol{\epsilon}_t^{\mathsf{a}})$ in the a-th copy into 2^{2K} classes according to the equivalence relation of $\Sigma\boldsymbol{e}_T^{\mathsf{a}}$ modulo stabilizers $\boldsymbol{a}_{\boldsymbol{v}}^{\mathsf{a}}$, i.e. $[\Sigma\boldsymbol{e}_T^{\mathsf{a}}] := \boldsymbol{\kappa}$.

With the newly introduced variables, one can express the n-th moment as the partition function of a stat-mech model, i.e. $\operatorname{tr}\rho_{Q\mathbf{M}}^n = \mathscr{Z}_n$,

$$\mathscr{Z}_n = \frac{1}{2^{K(n-1)}} \sum_{\boldsymbol{b}^1, \boldsymbol{\epsilon}^1} \Pr(\boldsymbol{b}^1, \boldsymbol{\epsilon}^1) \left(\sum_{\boldsymbol{\kappa}} \mathscr{Z}'(\boldsymbol{b}^1, \boldsymbol{\epsilon}^1, \boldsymbol{\kappa})\right)^{n-1}, \qquad (70)$$

where the prefactor $1/2^{K(n-1)}$ originates from the normalization of ρ_0, and

$$\mathscr{Z}'(\boldsymbol{b}^1, \boldsymbol{\epsilon}^1, \boldsymbol{\kappa})$$
$$= \sum_{\{\boldsymbol{e}_t\}} \Pr\left(\{\boldsymbol{b}_t^1 + \boldsymbol{e}_t, \epsilon_{t,i}^1 + \langle\langle \boldsymbol{a}_i, \Sigma\boldsymbol{e}_t\rangle\rangle\}\right) \delta_{[\Sigma\boldsymbol{e}_T]=\boldsymbol{\kappa}}. \qquad (71)$$

The partition function $\mathscr{Z}_n$ can be regarded as the disorder average of $n - 1$ copies of the disordered partition function. Specifically, we view $\boldsymbol{b}^1$ and $\boldsymbol{\epsilon}^1$ as disorder variables. The disordered stat-mech model involves binary variables $\boldsymbol{e}_t$ as the degrees of freedom. We introduce an additional label $\boldsymbol{\kappa}$ for later convenience; it is associated with the decoding class in the QEC problem. In the example of the toric code, the stat-mech model is a 3D RPIM, and $\boldsymbol{\kappa}$ indicates the defect insertions (i.e. π-flux lines) along non-contractible loops in two spatial directions. We note that the stat-mech model above and the model obtained

in the stabilizer expansion in Sec. IV A describe the same quantity and therefore are dual.

We remark that the disordered model in Eq. (70) has a gauge symmetry. The Boltzmann weight Pr in Eq. (71) is invariant under a local change of the disorder configuration

$$\boldsymbol{b}_t^1 \mapsto \boldsymbol{b}_t^1 + \boldsymbol{e}_t^1, \quad \delta\epsilon_{t,i}^1 \mapsto \delta\epsilon_{t,i}^1 + \langle\!\langle \boldsymbol{a}_i, \boldsymbol{e}_t^1 \rangle\!\rangle \qquad (72)$$

together with a redefinition of the degrees of freedom in the model[5],

$$\boldsymbol{e}_t \mapsto \boldsymbol{e}_t - \boldsymbol{e}_t^1. \qquad (73)$$

Hence, the partition function $\mathcal{Z}'$ is invariant under the shift in disorder configuration in Eq. (72), and its value $\mathcal{Z}'$ is fully determined by the erroneous syndromes $\boldsymbol{m}_t' := \boldsymbol{m}_t + \boldsymbol{\epsilon}_t$, true stabilizer eigenvalues $\boldsymbol{m}_T$ in the final state, and logical sectors $\boldsymbol{\kappa}$.

To understand the stat-mech model in the $n \to 1$ limit, we consider a specific information diagnostic, the coherent information, which according to Eq. (28) is related to the Shannon entropy of the distribution over decoding sectors in the limit $n \to 1$. In Appendix C, we show that, for each syndrome $(\boldsymbol{m}_t', \boldsymbol{m}_T)$, the partition function $\mathcal{Z}'$ determines the probability of each decoding sector $\boldsymbol{\kappa}$ in the ML decoder:

$$P(\boldsymbol{\kappa}|\boldsymbol{m}_t', \boldsymbol{m}_T) = \frac{\mathcal{Z}'(\boldsymbol{b}_0^1, \boldsymbol{\epsilon}_0^1, \boldsymbol{\kappa})}{\sum_{\boldsymbol{\kappa}} \mathcal{Z}'(\boldsymbol{b}_0^1, \boldsymbol{\epsilon}_0^1, \boldsymbol{\kappa})}, \qquad (74)$$

where $(\boldsymbol{b}_0^1, \boldsymbol{\epsilon}_0^1)$ is a reference error configuration that produces the syndrome $(\boldsymbol{m}_t', \boldsymbol{m}_T)$[6]. Therefore, the disordered model $\mathcal{Z}'$ governs the information diagnostics in the limit $n \to 1$ and is a direct generalization of the model previously obtained for the ML decoder in specific quantum codes [3, 15–20].

Importantly, we note that the disordered stat-mech model $\mathcal{Z}'$ exhibits an enlarged symmetry, which is a defining feature of the Nishimori line [14]. Specifically, the disorder average of $n-1$ copies of the partition function $\mathcal{Z}'$, i.e. $\mathcal{Z}_n$, exhibits a permutation symmetry S_n over n copies (as shown in Appendix C) [37, 38]. In our problem, the n-th moment $\mathcal{Z}_n = \mathrm{tr}\,\rho_{QM}^n$ always has an S_n symmetry which stems from the permutation symmetry over n identical copies of the density matrices ρ_{QM}.

We further note that, for perfect measurements, $\boldsymbol{e}_t = \boldsymbol{a}_{\boldsymbol{v}_t} + \overline{\boldsymbol{a}}_{\boldsymbol{\kappa}_t}$ given by a stabilizer and a logical operator, and $\delta\boldsymbol{\epsilon}_t^1 = 0$. The stat-mech model is decoupled in the

temporal direction, and the partition function $\mathcal{Z}'$ reduces to

$$\mathcal{Z}'(\boldsymbol{b}^1, \boldsymbol{\kappa}) = \prod_t \sum_{\boldsymbol{v}_t, \boldsymbol{\kappa}_t} \mathrm{Pr}\left(\boldsymbol{b}_t^1 + \overline{\boldsymbol{a}}_{\boldsymbol{\kappa}_t} + \boldsymbol{a}_{\boldsymbol{v}_t}\right) \delta_{\sum_t \boldsymbol{\kappa}_t = \boldsymbol{\kappa}}. \ (75)$$

The disordered stat-mech model for each time step is the same as the model in Ref. [19] derived for the ML decoder.

C. Duality between two expansions

Having derived stat-mech models based on the stabilizer expansion in Eq. (37) and error configuration expansion in Eq. (25), we here explicitly show that the two expansions are related by Fourier transformation.

Both expansions are possible because the set of stabilizers $g^{\boldsymbol{u}}$ and the set of syndrome projectors $\mathcal{P}_{\boldsymbol{m}}$ both comprise operator bases for the same space, which is the set of density matrices that are block diagonal with respect to syndromes. In particular, we note that $\mathcal{P}_{\boldsymbol{m}} = 2^{-I_0} \sum_{\boldsymbol{u}} (-1)^{\boldsymbol{m}\cdot\boldsymbol{u}} g^{\boldsymbol{u}}$ and $|\boldsymbol{m}\rangle\langle\boldsymbol{m}| = 2^{-I} \sum_{\boldsymbol{u}} (-1)^{\boldsymbol{m}\cdot\boldsymbol{u}} Z^{\boldsymbol{u}}$. Therefore, up to an overall constant,

$$\mathcal{P}_{\boldsymbol{m}} \bigotimes_{t=1}^T |\boldsymbol{m}_t'\rangle\langle\boldsymbol{m}_t'| = \sum_{\{\boldsymbol{u}_t\}} (-1)^{\sum_t \boldsymbol{\delta m}_t' \cdot \boldsymbol{u}_t} g^{\boldsymbol{u}_T} \prod_{t=1}^T Z_t^{\boldsymbol{u}_t - \boldsymbol{u}_{t-1}}$$
$$(76)$$

where $\boldsymbol{\delta m}_t' = \boldsymbol{m}_{t+1}' - \boldsymbol{m}_t'$ with $\boldsymbol{m}_{T+1}' := \boldsymbol{m}$ and $\boldsymbol{m}_0' := \boldsymbol{0}$. Eq. (76) may also be inverted to express stabilizers in terms of syndromes.

By comparing Eq. (25) and Eq. (37), and applying the change of basis rule from Eq. (76), we arrive at a relationship between the coefficients in the two expansions:

$$\mathrm{Pr}(\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}) = \sum_{\{\boldsymbol{u}_t\}} (-1)^{\sum_t \boldsymbol{\delta m}_t' \cdot \boldsymbol{u}_t} e^{-\mathcal{H}(\{\boldsymbol{u}_t\})}. \ (77)$$

We observe that this may be viewed as a $\mathbb{Z}_2$ Fourier transform, and observe that it relates partition functions with quenched disorder in the error configuration expansion with Boltzmann weights in the stabilizer expansion.

As we saw in Sec. IV B, $\mathrm{Pr}(\boldsymbol{m}_T, \{\boldsymbol{m}_t'\})$ may be written in terms of error configuration degrees of freedom. With this in mind, we can use Eq. (77) to quickly obtain two dualities.

First, we observe that

$$\mathrm{Pr}(\boldsymbol{0}, \{\boldsymbol{0}\}) = \sum_{\{\boldsymbol{u}_t\}} e^{-\mathcal{H}(\{\boldsymbol{u}_t\})}, \qquad (78)$$

which shows that the error configuration stat-mech model without disorder is dual to a single flavor of the stabilizer expansion stat-mech model.

Second, we observe that

$$\sum_{\boldsymbol{m}_T, \{\boldsymbol{m}_t'\}} \mathrm{Pr}(\boldsymbol{m}_T, \{\boldsymbol{m}_t'\})^n = \sum_{\{\boldsymbol{u}_t^a\}} e^{-\mathcal{H}^{(n)}(\{\boldsymbol{u}_t^a\})}, \qquad (79)$$

---

[5] In the 2D RBIM for decoding the toric code with perfect syndrome measurements, the gauge transformation is a spin flip in Eq. (73) combined with a shift in the random bond configuration as in Eq. (72).

[6] We note that the decoding sector $\boldsymbol{\kappa}$ depends on the choice of $(\boldsymbol{b}_0^1, \boldsymbol{\epsilon}_0^1)$; however, the probability of the most likely sector is independent of this choice.

since $\sum_{\boldsymbol{\delta m}} (-1)^{\sum_t \boldsymbol{\delta m}'_t \cdot \boldsymbol{u}_t}$ places a constraint on the n-th flavor of spins in terms of the other flavors, as we saw in Sec. IV A 2. In this way, we see that the structure of the duality arises from a generalized form of the Plancherel theorem.

D. Examples

Having developed the mapping for generic stabilizer codes, we here consider the surface code, repetition code, and the XZZX code as specific examples and explicitly derive the stat-mech models to study their decoding transitions. In what follows, we provide the summary of the results and leave the details in the appendices.

In Appendix D, we consider the surface code subject to single-qubit Pauli errors and noisy measurements. For Pauli-X (or Pauli-Z) error, the stat-mech model is a 3D $(n-1)$-flavor Ising model with nearest-neighbor couplings. In Appendix C, we verify that the model is the Kramers-Wannier dual of the 3D random plaquette Ising model in the limit $n \to 1$. For Pauli-Y decoherence, the stat-mech model has two-body couplings in the temporal direction and four-body plaquette couplings in the spatial direction; the model is in the same universality as the $(2+1)$D Xu-Moore model [39]. We note that the $(2+1)$D dynamics of the surface code can be re-formulated as virtual time evolution generated by sequentially measuring the qubits in a 3D cluster state introduced by Raussendorf, Bravyi and Harrington (RBH) [40]. In Appendix G we alternatively derive the stat-mech model for the $(2+1)$D surface code by considering the measurement dynamics of the noisy RBH state.

In Appendix H, we derive the stat-mech model for the d-dimensional repetition code subject to Pauli-X decoherence and noisy measurements. We show that the stat-mech model is a 2D $(n-1)$-flavor Ising model for the 1D repetition code, and the model is a 3D $(n-1)$-flavor $\mathbb{Z}_2$ gauge theory for the 2D repetition code.

In Appendix I, we consider the XZZX code subject to Pauli decoherence and noisy measurements. The resulting 3D $(n-1)$-flavor Ising model consists of two-body and four-body couplings, similar to the surface code. Notably, in the case of only Pauli-X (or Pauli-Z) decoherence, the model reduces to decoupled 2D Ising models.

V. DISCUSSION

In this work, we study the dynamics of quantum memories under local Pauli decoherence channels and repeated noisy syndrome measurements. We develop a stat-mech model for non-linear functions of the density matrix, which allows evaluating information diagnostics, such as coherent information and relative entropy.

Our model is dual to the stat-mech model previously obtained for the optimal decoding algorithm. This duality can be understood by considering two different expan-

sions of the same density matrix: the stabilizer expansion and the error configuration expansion. Writing the density matrix as a weighted sum of stabilizers leads to our model and writing the density matrix as a weighted sum of various errors applied to the initial state leads to the previously-obtained model.

Our dual model offers certain benefits. It exposes the impact of imperfect stabilizer measurements on information retention. The model consists of a layer of spins for each time step of the dynamics. After a single round of decoherence, quantities like the coherent information are determined by ordering in a single layer of spins. Perfect syndrome measurements ensure that successive layers of spins are completely decoupled from prior layers so that successive rounds of decoherence do not build on top of prior rounds. As the probability of readout noise increases, the layers of spins become coupled, making it possible for less information to be retained by the system over time. In the limit that readout noise is maximal, the layers are completely coupled together and it is as if no syndrome measurements occurred at all.

Our mapping also explains the necessity for repeating noisy measurements for successful error correction in the surface code even though the memory can in principle retain (maximum) coherent information even after $O(1)$ rounds of noisy measurements [3]. This originates from the clear distinction between stat-mech models for $\rho_{Q\mathbf{M}}$ and $\rho_{\mathbf{M}}$. In particular, error correction requires identifying the error configuration only from the noisy measurement record and therefore concerns the stat-mech model associated with $\rho_{\mathbf{M}}$. In the stabilizer expansion, this stat-mech model has a boundary magnetic field explicitly breaking the Ising symmetry and therefore cannot exhibit a paramagnetic phase for $O(1)$ rounds of noisy measurements. When the noisy measurements are repeated for $O(L)$ times, the boundary magnetic field does not affect the ordering in the bulk of the $(2+1)$D stat-mech model. Thus, the stat-mech model can be in the paramagnetic phase indicating successful error correction. On the other hand, the coherent information is associated with the stat-mech model for $\rho_{Q\mathbf{M}}$, which has open boundary conditions on the final layer of spins and can undergo a paramagnetic-to-ferromagnetic transition for any finite number of rounds of noisy measurements.

Our results open several directions for further study. First, we note that in the dual picture developed in this work, i.e. the stat-mech model in the stabilizer expansion, codes with maximum threshold correspond to spin models that fail to order at any finite temperature. As such, one may be able to use insights from statistical mechanics to reverse engineer QEC codes that have maximal thresholds even in the presence of noisy measurements.

Second, there are certain QEC codes whose many redundancies or other special properties enable single-shot QEC [41], in which single rounds of noisy measurements can be used for decoding. Our dual stat-mech model may provide a useful platform for studying single-shot stabilizer QEC codes. Note that the stat-mech model describ-

ing moments of $\rho_{\mathbf{M}}$ after a single round of measurements would be a spin model with an external field, which normally induces ferromagnetic order. On the other hand, as we discussed in Sec. IV, local redundancies in measured stabilizers translate to gauge (or higher form) symmetries of the stat-mech model, and it is known that higher-form symmetries can be emergent even if the symmetry is explicitly broken microscopically [42, 43]. This may provide a useful perspective explaining the essential ingredients for successful single-shot decoding.

Third, although this work focuses only on stabilizer codes, it may be possible to generalize our stat-mech model to other related QEC codes, e.g. subsystem codes [44] and dynamically generated codes [45], subject to Pauli errors. In both cases, if starting from the maximally mixed state in the code space, the density matrix of the system at any time step can be expressed as a classical mixture of Pauli operators and, as such, there may exist natural extensions of our stat-mech model in each of these settings.

Note added: After completing this manuscript, we became aware of a recent work [46], which obtains results related to ours in Sec. III C.

ACKNOWLEDGMENTS

[1] B. J. Brown, D. Loss, J. K. Pachos, C. N. Self, and J. R. Wootton, Reviews of Modern Physics **88**, 045005 (2016).

[2] B. M. Terhal, Reviews of Modern Physics **87**, 307 (2015).

[3] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Journal of Mathematical Physics **43**, 4452 (2002).

[4] R. Fan, Y. Bao, E. Altman, and A. Vishwanath, arXiv preprint arXiv:2301.05689 (2023).

[5] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, arXiv preprint arXiv:2301.05687 (2023).

[6] J. Y. Lee, C.-M. Jian, and C. Xu, arXiv preprint arXiv:2301.05238 (2023).

[7] Z. Wang, Z. Wu, and Z. Wang, arXiv preprint arXiv:2307.13758 (2023).

[8] K. Su, Z. Yang, and C.-M. Jian, arXiv preprint arXiv:2401.17359 (2024).

[9] A. Lyons, arXiv preprint arXiv:2403.03955 (2024).

[10] Y. Zhao and D. E. Liu, arXiv preprint arXiv:2312.16991 (2023).

[11] L. Colmenarez, Z.-M. Huang, S. Diehl, and M. Müller, arXiv preprint arXiv:2312.06664 (2023).

[12] J. Y. Lee, arXiv preprint arXiv:2402.16937 (2024).

[13] R. Sohal and A. Prem, arXiv preprint arXiv:2403.13879 (2024).

[14] H. Nishimori, Progress of Theoretical Physics **66**, 1169 (1981).

[15] C. Wang, J. Harrington, and J. Preskill, Annals Phys. **303**, 31 (2003), arXiv:quant-ph/0207088.

[16] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, Phys. Rev. Lett. **103**, 090501 (2009), arXiv:0902.4845 [cond-mat.dis-nn].

[17] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, Phys. Rev. X **2**, 021004 (2012), arXiv:1202.1852 [quant-ph].

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

[19] C. T. Chubb and S. T. Flammia, Annales de l'Institut Henri Poincaré D **8**, 269 (2021).

[20] H. Song, J. Schönmeier-Kromer, K. Liu, O. Viyuela, L. Pollet, and M. A. Martin-Delgado, Phys. Rev. Lett. **129**, 230502 (2022), arXiv:2112.05122 [quant-ph].

[21] F. Venn, J. Behrends, and B. Béri, (2022), arXiv:2211.00655 [quant-ph].

[22] J. Behrends, F. Venn, and B. Béri, (2022), arXiv:2212.08084 [quant-ph].

[23] D. Gottesman, *Stabilizer codes and quantum error correction* (California Institute of Technology, 1997).

[24] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996), arXiv:quant-ph/9604022.

[25] B. Schumacher and M. D. Westmoreland, Quantum Information Processing **1**, 5 (2002), arXiv:quant-ph/0112106 [quant-ph].

[26] J. Preskill, Caltech Lecture Notes **7**, 1 (1999).

[27] M. M. Wilde, *Quantum information theory* (Cambridge university press, 2013).

[28] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, Physical Review Letters **125**, 030505 (2020).

[29] E. H. Lieb and M. B. Ruskai, Les rencontres physiciens-mathématiciens de Strasbourg-RCP25 **19**, 36 (1973).
[30] H. Araki, Publications of the Research Institute for Mathematical Sciences **11**, 809 (1975).
[31] G. Lindblad, Communications in Mathematical Physics **40**, 147 (1975).
[32] A. Uhlmann, Communications in Mathematical Physics **54**, 21 (1977).
[33] Y. Bao, S. Choi, and E. Altman, Physical Review B **101**, 104301 (2020).
[34] S. Bravyi, M. Suchara, and A. Vargo, Physical Review A **90**, 032326 (2014).
[35] P. Iyer and D. Poulin, IEEE Transactions on Information Theory **61**, 5209 (2015).
[36] Y. Li, N. O'Dea, and V. Khemani, arXiv preprint arXiv:2406.15757 (2024).
[37] P. Le Doussal and A. B. Harris, Physical review letters **61**, 625 (1988).
[38] I. A. Gruzberg, N. Read, and A. W. Ludwig, Physical Review B **63**, 104422 (2001).
[39] C. Xu and J. E. Moore, Physical review letters **93**, 047003 (2004).
[40] R. Raussendorf, S. Bravyi, and J. Harrington, Physical Review A **71**, 062313 (2005).
[41] H. Bombín, Physical Review X **5**, 031043 (2015).
[42] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, Journal of High Energy Physics **2015**, 1 (2015).
[43] J. McGreevy, Annual Review of Condensed Matter Physics **14**, 57 (2023).
[44] D. Poulin, Physical review letters **95**, 230504 (2005).
[45] M. B. Hastings and J. Haah, Quantum **5**, 564 (2021).
[46] R. Niwa and J. Y. Lee, arXiv preprint arXiv:2407.02564 (2024).
[47] E. Fradkin, B. A. Huberman, and S. H. Shenker, Physical Review B **18**, 4789 (1978).
[48] A. Y. Kitaev, in *Quantum communication, computing, and measurement* (Springer, 1997) pp. 181–188.
[49] A. Y. Kitaev, Annals of physics **303**, 2 (2003).
[50] G. Semeghini *et al.*, Science **374**, abi8794 (2021), arXiv:2104.04119 [quant-ph].
[51] D. Bluvstein *et al.*, Nature **604**, 451 (2022), arXiv:2112.03923 [quant-ph].
[52] K. J. Satzinger *et al.*, Science **374**, abi8378 (2021), arXiv:2104.01180 [quant-ph].
[53] R. Acharya *et al.* (Google Quantum AI), (2022), arXiv:2207.06431 [quant-ph].
[54] T. I. Andersen *et al.*, (2022), arXiv:2210.10255 [quant-ph].
[55] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Nature , 1 (2023).
[56] D. K. Tuckett, A. S. Darmawan, C. T. Chubb, S. Bravyi, S. D. Bartlett, and S. T. Flammia, Physical Review X **9**, 041031 (2019).
[57] H. Bombin and M. A. Martin-Delgado, Journal of mathematical physics **48** (2007).
[58] M. N. Barber, R. Pearson, D. Toussaint, and J. L. Richardson, Physical Review B **32**, 1720 (1985).
[59] A. M. Ferrenberg and D. Landau, Physical Review B **44**, 5081 (1991).
[60] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, Nature communications **12**, 2172 (2021).
[61] A. Bolt, G. Duclos-Cianci, D. Poulin, and T. Stace, Physical review letters **117**, 070501 (2016).
[62] M. Van den Nest, J. Dehaene, and B. De Moor, Physical Review A **69**, 022316 (2004).
[63] R. Raussendorf and H. J. Briegel, arXiv preprint quant-ph/0010033 (2000).
[64] R. Raussendorf, International Journal of Quantum Information **7**, 1053 (2009).

Appendix A: Generalization to weak measurements

In this work, we have focused on imperfect measurements where the imperfection is caused by a readout error that changes a measurement result. Weak measurements are another natural source of imperfections, where the unitary $U_{Q \to QM}$ fails to perfectly correlate the system and measurement ancilla, even before any readout errors occur. This interaction can be described by the same form of $U_{Q \to QM}$ in Eq. (7) but with $\mathcal{P}_{0,1}$ modified such that

$$\mathcal{P}_{0,1} = \frac{1 \pm \lambda g}{\sqrt{2(1+\lambda^2)}} \tag{A1}$$

where $0 \leq \lambda \leq 1$. When $\lambda = 1$ the operators $\mathcal{P}_{0,1}$ are unchanged from the rest of the text, and the measurement becomes weaker the smaller λ becomes.

As long as the initial state is in the code space and is subject only to Pauli errors and imperfect syndrome measurements, the state ρ at any later time may be written as a weighted sum over stabilizers. Consequently, the state always commutes with $\mathcal{P}_{0,1}$. Noting also that $\mathcal{P}_0$ and $\mathcal{P}_1$ commute with each other, and using the shorthand U for $U_{Q \to Q\mathbf{M}}$, we can write

$$U\rho U^\dagger = \rho[\mathcal{P}_0^2 \otimes |0\rangle\langle 0|_M + \mathcal{P}_1^2 \otimes |1\rangle\langle 1|_M \\ + \mathcal{P}_0\mathcal{P}_1 \otimes (|0\rangle\langle 1|_M + |1\rangle\langle 0|_M)]. \tag{A2}$$

We note that

$$\mathcal{P}_{0,1}^2 = \frac{1}{2}\left(1 \pm \frac{2\lambda g}{1+\lambda^2}\right). \tag{A3}$$

It is useful to write $|0\rangle\langle 0|_M = (1 + Z_M)/2$, $|1\rangle\langle 1|_M = (1 - Z_M)/2$, and $|0\rangle\langle 1|_M + |1\rangle\langle 0|_M = X_M$, because the behavior of Z_M and X_M under readout errors is straightforward. In particular,

$$\mathcal{N}_M[Z_M] = (1 - 2q)Z_M \tag{A4}$$
$$\mathcal{N}_M[X_M] = X_M. \tag{A5}$$

It is also useful to define weights $\mathcal{W}_q = (1 - 2q)$ and $\mathcal{W}_\lambda = 2\lambda/(1+\lambda^2)$. Then,

$$\mathcal{N}_M[U\rho U^\dagger] = \frac{1}{4}\rho[(1 + \mathcal{W}_\lambda g)(1 + \mathcal{W}_q Z_M) \tag{A6}$$
$$+ (1 - \mathcal{W}_\lambda g)(1 - \mathcal{W}_q Z_M) \tag{A7}$$
$$+ \mathcal{P}_0\mathcal{P}_1 X_M]. \tag{A8}$$

In Sec. II, the $\mathcal{P}_0\mathcal{P}_1 X_M$ term was not present because $\mathcal{P}_0\mathcal{P}_1 = 0$ when $\lambda = 1$. As a result, it was not necessary

to explicitly consider the impact of the actual projective measurement of the ancilla in the Pauli-Z basis. Here, it is necessary to consider this measurement, leading us to modify the imperfect measurement channel $\mathcal{M}$:

$$\mathcal{M}[\rho] = \left(\frac{1+Z_M}{2}\right)\mathcal{N}_M[U\rho U^\dagger]\left(\frac{1+Z_M}{2}\right)$$
$$+ \left(\frac{1-Z_M}{2}\right)\mathcal{N}_M[U\rho U^\dagger]\left(\frac{1-Z_M}{2}\right) \quad (A9)$$

which annihilates the $\mathcal{P}_0\mathcal{P}_1 X_M$ term. We conclude that

$$\mathcal{M}[\rho] = \frac{1}{2}\rho(1 + \mathcal{W}_\lambda \mathcal{W}_q g Z_M). \quad (A10)$$

Analogous work in Sec. IV A 1 yields the same result without the factor of $\mathcal{W}_\lambda$. Since $\mathcal{W}_\lambda$ is always positive, the combined weight $\mathcal{W}_\lambda \mathcal{W}_q$ is positive as long as $0 \le q \le \frac{1}{2}$. In Sec. IV A 1 these weights and their logarithms are natural objects to consider, allowing us to interpret the impact of weak measurements as modifying

$$\mu_q \to \mu_q - \log\frac{2\lambda}{1+\lambda^2} \quad (A11)$$

where $\mu_q = -\log(1 - 2q)$. We can also view this as modifying

$$q \to q\left(\frac{2\lambda}{1+\lambda^2}\right) + \frac{1}{2}\frac{(1-\lambda)^2}{1+\lambda^2} \quad (A12)$$

which lies between 0 and $\frac{1}{2}$ as long as q does.

In summary, allowing weak measurements in our dynamics is equivalent to modifying the rate of readout errors. However, we note that the situation would be significantly more complicated in dynamics with coherent errors.

Appendix B: Bound on the optimal decoding infidelity

Here, we provide a proof of the inequality

$$\overline{\Delta} \le H(\boldsymbol{\kappa}|X) \quad (B1)$$

where we use the shorthand notation $X := (\boldsymbol{m}_T, \{\boldsymbol{m}'_t\})$, and

$$\overline{\Delta} = \sum_X \Pr(X)(1 - \max_{\boldsymbol{\kappa}} \Pr(\boldsymbol{\kappa}|X)). \quad (B2)$$

For fixed X, $H_\infty(\Pr(\boldsymbol{\kappa}|X)) = -\log\max_{\boldsymbol{\kappa}} \Pr(\boldsymbol{\kappa}|X)$. It follows that

$$\overline{\Delta} = \sum_X \Pr(X)(1 - e^{-H_\infty(\Pr(\boldsymbol{\kappa}|X))}). \quad (B3)$$

Since $1 - e^{-x}$ is concave, the linear approximation of the function at any point is an upper bound. Consequently,

$$\overline{\Delta} \le \sum_X \Pr(X)H_\infty(\Pr(\boldsymbol{\kappa}|X)) \quad (B4)$$

and, using Rényi entropy inequalities,

$$\overline{\Delta} \le \sum_X \Pr(X)H(\Pr(\boldsymbol{\kappa}|X)) \quad (B5)$$

where $H(\Pr(\boldsymbol{\kappa}|X)) = -\sum_{\boldsymbol{\kappa}} \Pr(\boldsymbol{\kappa}|X)\log\Pr(\boldsymbol{\kappa}|X)$. It follows that

$$\overline{\Delta} \le -\sum_{\boldsymbol{\kappa},X} \Pr(X)\Pr(\boldsymbol{\kappa}|X)\log\Pr(\boldsymbol{\kappa}|X) \quad (B6)$$

$$= -\sum_{\boldsymbol{\kappa},X} \Pr(\boldsymbol{\kappa},X)\log\Pr(\boldsymbol{\kappa}|X) \quad (B7)$$

$$= H(\boldsymbol{\kappa}|X). \quad (B8)$$

Appendix C: The $n \to 1$ limit of the stat-mech model in the error configuration expansion

We here show that the disordered stat-mech model $\mathcal{Z}'$ governs the information theoretical diagnostics in the limit $n \to 1$. We consider coherent information as an example.

We first take the error configuration expansion of the n-th moment $\operatorname{tr}\rho_{Q\mathbf{M}R}^n$

$$\operatorname{tr}\rho_{Q\mathbf{M}R}^n = \sum_{\boldsymbol{b}^1,\boldsymbol{\epsilon}^1} \Pr(\boldsymbol{b}^1,\boldsymbol{\epsilon}^1)\mathcal{Z}'(\boldsymbol{b}^1,\boldsymbol{\epsilon}^1,\mathbf{0})^{n-1}. \quad (C1)$$

Here, the term in the error configuration expansion of $\rho_{Q\mathbf{M}R}$ contributes to $\operatorname{tr}\rho_{Q\mathbf{M}R}^n$ only if the cumulative error $\Sigma\boldsymbol{e}_T^\mathrm{a} = \Sigma\boldsymbol{b}_T^\mathrm{a} - \Sigma\boldsymbol{b}_T^1$ is given by a stabilizer, i.e. $\boldsymbol{\kappa} = [\Sigma\boldsymbol{e}_T^\mathrm{a}] = \mathbf{0}$.

Next, knowing that the partition function $\mathcal{Z}'$ only depends on $(\boldsymbol{m}'_t, \boldsymbol{m}_T)$, we can express the n-th moment in terms of the probability of error strings belonging to different sectors, i.e.

$$\operatorname{tr}\rho_{Q\mathbf{M}}^n = \frac{1}{2^{K(n-1)}}\sum_{\boldsymbol{m}_T,\boldsymbol{m}'_t}\left(\sum_{\boldsymbol{\kappa}}\Pr(\boldsymbol{\kappa},\boldsymbol{m}_T,\boldsymbol{m}'_t)\right)^n,$$

$$\operatorname{tr}\rho_{Q\mathbf{M}R}^n = \sum_{\boldsymbol{m}_T,\boldsymbol{m}'_t}\sum_{\boldsymbol{\kappa}}\Pr(\boldsymbol{\kappa},\boldsymbol{m}_T,\boldsymbol{m}'_t)^n, \quad (C2)$$

where

$$\Pr(\boldsymbol{\kappa},\boldsymbol{m}_T,\boldsymbol{m}'_t) := \mathcal{Z}'(\boldsymbol{b}_0^1,\boldsymbol{\epsilon}_0^1,\boldsymbol{\kappa}). \quad (C3)$$

Here, $\boldsymbol{b}_0^1,\boldsymbol{\epsilon}_0^1$ are a reference error configuration, and we have the relation $\mathcal{Z}'(\boldsymbol{b}^1,\boldsymbol{\epsilon}^1,\mathbf{0}) = \Pr(\boldsymbol{\kappa},\boldsymbol{m}_T,\boldsymbol{m}'_t)$ for $[\Sigma\boldsymbol{b}^1 - \Sigma\boldsymbol{b}_0^1] = \boldsymbol{\kappa}$. This expression explicitly shows the enlarged S_n permutation symmetry in the n-th moment $\mathcal{Z}_n$.

In the limit $n \to 1$, the coherent information reduces to $I_c = K\log 2 - H(\boldsymbol{\kappa}|\boldsymbol{m}_T,\boldsymbol{m}'_t)$ in agreement with Eq. (28), where $H(\boldsymbol{\kappa}|\boldsymbol{m}_T,\boldsymbol{m}'_t)$ is the Shannon entropy of the conditional distribution

$$\Pr(\boldsymbol{\kappa}|\boldsymbol{m}_T,\boldsymbol{m}'_t) = \frac{\mathcal{Z}'(\boldsymbol{b}_0^1,\boldsymbol{\epsilon}_0^1,\boldsymbol{\kappa})}{\sum_{\boldsymbol{\kappa}}\mathcal{Z}'(\boldsymbol{b}_0^1,\boldsymbol{\epsilon}_0^1,\boldsymbol{\kappa})}. \quad (C4)$$

This calculation shows that the partition function $\mathcal{Z}'$ determines the coherent information I_c. Moreover, $\mathcal{Z}'$ is proportional to the probability of each decoding sector and therefore is exactly the model obtained for the ML decoder [3].

Appendix D: Decoding transitions in the surface code

The surface code [48, 49] is a promising quantum memory that can retain quantum information up to a finite threshold of decoherence and has been realized in leading experimental platforms [50–55]. The information dynamics in the decohered surface code was previously studied in the context of optimal decoding algorithms [3]. It was shown that the decoding transition based on a single round of measurements (with no readout errors) maps to a ferromagnetic transition in the 2D random bond Ising model, and the transition in the case of repeated imperfect measurements maps to a deconfinement transition in the 3D random plaquette Ising model (RPIM). Here, we apply the stat-mech model developed in Sec. IV to the surface code and provide a dual perspective on the decoding transition in the case of repeated stabilizer measurements, generalizing the result in Ref. [4].

As we explain below, the stat-mech model derived using the stabilizer expansion in Sec. IV A exhibits a $\mathbb{Z}_2$ global symmetry and undergoes a ferromagnetic transition when tuning the decoherence rate, dual to the deconfinement transition in the RPIM. The decoherence rate tunes the temperature in the stat-mech model; the low-temperature ferromagnetic ordering corresponds to the information loss at a high decoherence rate. Our dual picture makes it easier to understand situations with a maximum threshold, which manifest as stat-mech models exhibiting no finite-temperature ferromagnetic ordering.

In the rest of this section, we first introduce the surface code and develop the stat-mech model in the dual picture. We then consider the surface code subject to physical Pauli-X/Z errors under repeated imperfect measurements. We obtain a 3D $(n-1)$-flavor Ising model for the Rényi-n quantities, which is dual to the previously known 3D RPIM in the limit $n \to 1$ [3] (shown in Appendix E). Next, we consider the surface code subject to Pauli-Y errors, which is shown to have the maximum error threshold ($p_c = 0.5$) based on a single round of perfect measurements [56]. Using our stat-mech mapping, we show that unbiased measurement errors result in a sub-maximum threshold ($p_c < 0.5$).

1. Surface code

A surface code is a stabilizer code associated with any cellulation of a two-dimensional manifold. We focus on the case of closed manifolds. Our results can be generalized to the planar surface code defined on an open
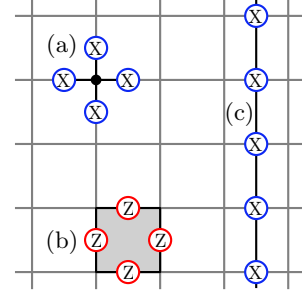

Figure 2. Surface code on the square lattice with periodic boundary conditions. The code has stabilizers A_v associated with vertices (a) and B_f associated with plaquettes (b). The logical operator is a string operator along a non-contractible loop (c).

manifold with suitable boundary conditions. The cellulation defines a graph (2-complex) that consists of vertices (0-cells), edges (1-cells), and faces (2-cells). The surface code consists of qubits on the edges ℓ of the graph and two types of stabilizers associated with vertices v and faces f:

$$A_v = \prod_{\ell \in \mathrm{star}(v)} X_\ell, \quad B_f = \prod_{\ell \in \mathrm{boundary}(f)} Z_\ell, \qquad \text{(D1)}$$

where $\mathrm{star}(v)$ and $\mathrm{boundary}(f)$ include all the edges emanated from the vertex v and on the boundary of the face f, respectively.

The logical space of the surface code depends on the homology of the manifold. Specifically, the stabilizer B_f (A_v) is a product of Pauli operators along a boundary (co-boundary) on the graph. The operators that commute with all the stabilizers are thus products of Pauli-Z(X) along cycles (co-cycles). It follows that logical-Z(X) operators, which are not products of stabilizers, are given by the cycles (co-cycles) that are not boundaries (co-boundaries) and are thus determined by the homology (co-homology) group of the graph. For a closed 2D manifold of genus G, the corresponding logical space can encode $K = 2G$ qubits. See Ref. [57] for details.

A widely studied example of the surface code is the so-called toric code defined on a two-dimensional square lattice with periodic boundary conditions (topologically equivalent to a torus) [48, 49]. The code space is four-fold degenerate and can encode two logical qubits. The logical-Z(X) operators are string operators along two non-contractible loops on the direct (dual) lattice on the torus. Examples of these check operators and logical operators are illustrated in Fig. 2. Our results in this section apply to general surface codes, but we use the toric code as a helpful example throughout.

We remark that the (2+1)D dynamics of the toric code under repeated syndrome measurements can be equivalently formulated as sequentially measuring the 3D Raussendorf-Bravyi-Harrington (RBH) state [40]. We

	Coherent information $I_c(R\rangle Q\mathbf{M})$	Quantum relative entropy $D_s(\rho_{Q\mathbf{M}}||\rho_{Q\mathbf{M},s})$	KL divergence $D_{\mathrm{KL},s}(\rho_{\mathbf{M}}||\rho_{\mathbf{M},s})$
Boundary field	×	×	✓
Finite threshold at $T = O(1)$	✓	✓	×
Finite threshold at $T = O(L)$	✓	✓	✓

Table I. Properties of the stat-mech model for information diagnostics in the decohered surface code under repeated stabilizer measurements.

detail this formulation and an alternative way to derive the stat-mech model in Appendix G.

2. Stat-mech model of the surface code

The stat-mech model for the surface code under repeated measurements involves Ising spins at each time step t associated with stabilizer measurements of A_v and B_f. Thus, the Ising spins are defined on the vertices of the graph $\mathcal{G}$ and the dual graph $\mathcal{G}^*$ denoted by $\sigma_{r,t}$ and $\sigma_{\bar{r},t}$, respectively. In the stabilizer expansion in Eq. (42), the Boltzmann weight is given by an effective Hamiltonian

$$\mathcal{H} = \sum_{t,\ell} -J_z \sigma_{r,t}\sigma_{r',t} - J_x \sigma_{\bar{r},t}\sigma_{\overline{r'},t} - J_y \sigma_{r,t}\sigma_{r',t}\sigma_{\bar{r},t}\sigma_{\overline{r'},t}$$
$$- \sum_{t,r} J_v \sigma_{r,t}\sigma_{r,t+1} - \sum_{t,\bar{r}} J_f \sigma_{\bar{r},t}\sigma_{\bar{r},t+1}, \qquad (D2)$$

where (r, r') is a pair of vertices connected by the edge ℓ, $(\bar{r}, \overline{r'})$ is a pair of faces (vertices on the dual graph) connected by the dual edge ℓ^*, and the couplings are given by Eq. (34) along with

$$J_v = -\frac{1}{2}\log(1 - 2q_v), \qquad (D3)$$

$$J_f = -\frac{1}{2}\log(1 - 2q_f). \qquad (D4)$$

Here, q_v and q_f denote the readout error rates for measuring A_v and B_f, respectively.

We can now write down the n-th moment as the partition function of a stat-mech model of $n - 1$ flavors of Ising spins, $\mathrm{tr}\,\rho_{Q\mathbf{M}}^n = \mathcal{Z}_n$. The symmetry of the stat-mech model is determined by the redundancies of the check operators. Here, the product of all star operators as well as the product of all plaquette operators is the identity, leading to the redundancy group $\mathcal{R} \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. Hence, the stat-mech model for the n-th moment exhibits a global symmetry of $\mathbb{G}^{(n)} = (\mathcal{R}^n \rtimes S_n)/\mathcal{R}$. When increasing the decoherence rate above the threshold, we expect a transition from a symmetric paramagnetic phase to a ferromagnetic phase that breaks the $\mathbb{G}^{(n)}$ symmetry. In what follows, we formulate the information diagnostics as observables in the stat-mech model and show they indeed detect the transition.

Coherent information.— Using Eq. (45), the n-th moment $\mathrm{tr}\,\rho_{RQ\mathbf{M}}^n$ can be expressed as

$$\mathrm{tr}\,\rho_{RQ\mathbf{M}}^n = \frac{1}{2^{2(n-1)}}\sum_{\{\boldsymbol{\kappa}^{\mathsf{a}}\}}\mathcal{Z}_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\}), \qquad (D5)$$

where $\boldsymbol{\kappa}^{\mathsf{a}}$ for $\mathsf{a} = 1,\dots,n-1$ is a 4-component binary vector; the first (last) two components label the insertion of two-dimensional non-contractible defects in xt and yt planes associated with logical-X(Z) operators, respectively.

We thus can write the n-th coherent information as

$$I_c^{(n)} = \frac{1}{n-1}\log\left(\sum_{\{\boldsymbol{\kappa}^{\mathsf{a}}\}}\frac{\mathcal{Z}_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\})}{\mathcal{Z}_n}\right) - 2\log 2. \qquad (D6)$$

We note that $-\log(\mathcal{Z}_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\})/\mathcal{Z}_n) = \Delta F_n(\{\boldsymbol{\kappa}^{\mathsf{a}}\})$ is the excess free energy of inserting the defects denoted by $\boldsymbol{\kappa}^{\mathsf{a}}$. In the paramagnetic phase, the defect costs zero free energy in the thermodynamic limit, so all the $2^{4(n-1)}$ terms in the summation are unity, thus, $I_c^{(n)} = 2\log 2$. In contrast, in the ferromagnetic phase of σ^{a} associated with Pauli-Z strings (assuming only Pauli-X errors and no Pauli-Z errors), the defect associated with logical-Z operators costs a free energy $O(LT)$ that increases with the system size. In this case, only $2^{2(n-1)}$ terms in the summation are unity, and we have $I_c^{(n)} = 0$. Hence, the decoding transition in the Rényi-n coherent information can be understood as the ordering transition in the 3D stat-mech model.

Relative entropy.— Here, we find the Rényi-n relative entropy in the $(n-1)$-flavor Ising model. We consider the relative entropy between $\rho_{Q\mathbf{M}}$ and $\rho_{Q\mathbf{M},m}$ where the latter state is evolved from the excited state $\rho_m = w_m(\gamma)\rho_0 w_m^\dagger(\gamma)$ with $w_m(\gamma) = \prod_{\ell \in \gamma} X_\ell$ creating m anyons at the endpoints r_1 and r_2 of γ.

Then, according to Eq. (56), the n-th Rényi relative entropy maps to the spin-spin correlation function on the bottom boundary of the stat-mech model

$$D_m^{(n)} = -\frac{1}{n-1}\log\left\langle\sigma_{0,r_1}^1\sigma_{0,r_2}^1\right\rangle. \qquad (D7)$$

Hence, the relative entropy detects the ferromagnetic transition in the $(n-1)$-flavor Ising model; it diverges in the paramagnetic phase and takes a finite value in the ferromagnetic phase.

Similarly, the KL divergence $D_{\mathrm{KL},m}^{(n)}$ maps to the correlation function in the stat-mech model with a boundary

field. When the number of time steps $T = O(L)$, the model can undergo a ferromagnetic phase transition in the bulk, which is detected by the KL divergence.

3. Threshold of Pauli-X/Z error

The first example we consider is the surface code subject to Pauli-Z (or -X) errors on the physical qubits, i.e. $p_x = p_y = 0$ (or $p_y = p_z = 0$). We focus on the case of Pauli-Z errors. The analysis for Pauli-X is similar.

The Ising spins on the direct and dual graph are decoupled in the stat-mech model, and the only non-vanishing couplings are $J_z = -(1/2)\log(1-2p_z)$, J_v, and J_f. We now consider the n-th moment $\operatorname{tr}\rho_{\mathrm{QM}}^n$ and write the effective Hamiltonian of the stat-mech model.

The stat-mech model for the spins on the dual graph $\mathcal{G}^*$ is decoupled one-dimensional Ising models with ferromagnetic couplings between neighboring sites in the temporal direction, which are always disordered for $q_f < 0.5$. This is consistent with the fact that Pauli-Z errors do not flip the eigenvalues of B_f stabilizers, and the classical information encoded in the logical Pauli-Z is always retained.

The stat-mech model for the spins on the direct graph $\mathcal{G}$ is a 3D multiple-flavor Ising model

$$
\mathcal{H}_Z^{(n)} = \sum_{\ell,t} -J_z \sum_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r',t}^{\mathsf{a}} - J_z \prod_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r',t}^{\mathsf{a}}
$$
$$
+ \sum_{r,t} -J_v \sum_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r,t+1}^{\mathsf{a}} - J_v \prod_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r,t+1}^{\mathsf{a}}, \quad \text{(D8)}
$$

where $\mathsf{a} = 1, 2, \ldots, n-1$ label the Ising spins in the a-th replica copy. The model is dual to the RPIM in the limit $n \to 1$ as shown in Appendix E.

The model undergoes a paramagnetic-to-ferromagnetic phase transition when increasing the error rate. In the case of $n = 2$, the stat-mech model reduces to a 3D Ising model with couplings $2J_z$ and $2J_v$. In the isotropic limit $p = q$ (i.e. $J_z = J_v$), the transition occurs at $2\mu_{z,c}^{(2)} = 0.221$ [58, 59], corresponding to $p_c^{(2)} = q_c^{(2)} = 0.099$. In the limit $n \to \infty$, the inter-replica interaction is negligible, giving rise to a 3D Ising model with couplings J_z and J_v. In the isotropic limit $p = q$, the transition occurs at $p_c^{(\infty)} = q_c^{(\infty)} = 0.179$.

4. Threshold of Pauli-Y error

The next example we consider is the surface code with Pauli-Y errors $(p_x = p_z = 0)$. The Ising spins on the direct and dual graph are coupled in the stat-mech model. The couplings are given by $J_x = J_z = 0$ and $J_y = -(1/2)\log(1-2p_y)$. We remark that, in this case, the stat-mech model exhibits subsystem symmetries in each flavor. In what follows, we analyze the threshold for different types of measurement errors.

a. Perfect measurement

We start by considering the case with perfect measurements. Here, the 3D multi-flavor Ising model consists of decoupled 2D multi-flavor Ising models in the temporal direction. The 2D model has a Hamiltonian

$$
\mathcal{H}_{Y,0}^{(n)} = -J_y \sum_{\ell} \left(\sum_{\mathsf{a}=1}^{n-1} \sigma_r^{\mathsf{a}}\sigma_{r'}^{\mathsf{a}}\sigma_{\bar{r}}^{\mathsf{a}}\sigma_{\bar{r}'}^{\mathsf{a}} + \prod_{\mathsf{a}=1}^{n-1} \sigma_r^{\mathsf{a}}\sigma_{r'}^{\mathsf{a}}\sigma_{\bar{r}}^{\mathsf{a}}\sigma_{\bar{r}'}^{\mathsf{a}} \right). \quad \text{(D9)}
$$

The model is a plaquette Ising model in 2D, which exhibits no finite temperature order for any replica index n. The absence of finite temperature order indicates a maximum threshold $p_c^{(n)} = 0.5$. This is consistent with $p_c = 0.5$ in the replica limit [56].

We remark that the maximum threshold in the case of Pauli-Y errors is relevant for realizing the surface code in experimental platforms with biased noise, i.e. when Pauli errors along a specific direction dominate. One can then tailor the surface code such that the biased noise is in the Pauli-Y direction and allow the code to achieve a high threshold of encoding quantum information [56]. With the same motivation, the XZZX code, related to the surface code by local rotations, has been proposed to encode information up to a maximum threshold of Pauli-X, Y, or Z errors [60].

b. Noisy measurement

We now consider the surface code subject to Pauli-Y errors and noisy measurements with unbiased readout errors. In this case, in addition to the four-body Ising couplings among spins at the same time step, Ising spins on consecutive time steps are also coupled by two-body Ising terms. The result is an Ising spin model on a cubic lattice (tilted from the original lattice by $\pi/4$) that contains $2L^2$ spins in each one of the T time steps (illustrated in Fig. 3(a)). The n-th moment maps the partition function with the Hamiltonian

$$
\mathcal{H}_Y^{(n)} =
$$
$$
-J_y \sum_{\ell,t} \sum_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r',t}^{\mathsf{a}}\sigma_{\bar{r},t}^{\mathsf{a}}\sigma_{\bar{r}',t}^{\mathsf{a}} + \prod_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r',t}^{\mathsf{a}}\sigma_{\bar{r},t}^{\mathsf{a}}\sigma_{\bar{r}',t}^{\mathsf{a}}
$$
$$
-J \sum_{r,t} \sum_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r,t+1}^{\mathsf{a}} + \prod_{\mathsf{a}=1}^{n-1} \sigma_{r,t}^{\mathsf{a}}\sigma_{r,t+1}^{\mathsf{a}}
$$
$$
-J \sum_{\bar{r},t} \sum_{\mathsf{a}=1}^{n-1} \sigma_{\bar{r},t}^{\mathsf{a}}\sigma_{\bar{r},t+1}^{\mathsf{a}} + \prod_{\mathsf{a}=1}^{n-1} \sigma_{\bar{r},t}^{\mathsf{a}}\sigma_{\bar{r},t+1}^{\mathsf{a}}. \quad \text{(D10)}
$$

where $J := J_v = J_f = -(1/2)\log(1-2q)$, and q is the unbiased readout error rate.

The phase boundary of the model can be analytically determined for certain replica indices. For $n = 2$, the model is self-dual [39]. Assuming only two phases of
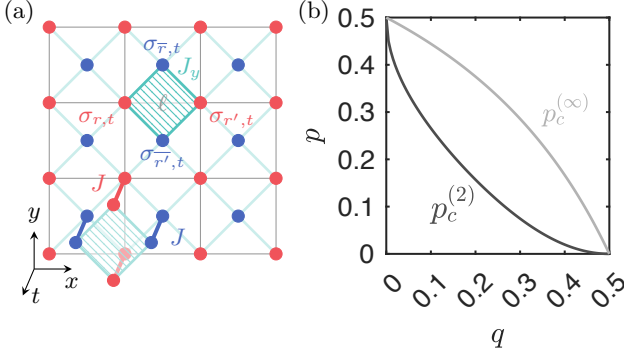
Figure 3. Critical error rates $(p_c^{(n)}, q_c^{(n)})$ in the $(n-1)$-flavor Ising model for the toric code with Pauli-Y errors and syndrome measurement errors. The physical and measurement error rates are denoted by p and q, respectively. The critical points are determined analytically for $n = 2$ (dark) and $n \to \infty$ (light) based on the self-duality of the model.

the model, a ferromagnetic and a paramagnetic phase, the transition occurs at the self-dual line satisfying $\sinh(4J_y)\sinh(4J) = 1$, i.e.

$$p_c^{(2)} = \frac{1}{2}\left(1 - \sqrt{\frac{1 - \left(1 - 2q_c^{(2)}\right)^2}{1 + \left(1 - 2q_c^{(2)}\right)^2}}\right). \tag{D11}$$

In the limit $n \to \infty$, neglecting the inter-replica interaction, the self-dual line is given by $\sinh(2J_y)\sinh(2J) = 1$, i.e.

$$p_c^{(\infty)} = 1 - \frac{1}{2 - 2q_c^{(\infty)}}. \tag{D12}$$

The critical points for $n = 2$ and $n \to \infty$ are illustrated in Fig. 3(b). Away from the limit of perfect measurement $q = 0$, the stat-mech model exhibits a finite temperature ferromagnetic phase, indicating a threshold that is less than the maximum value, i.e. $p_c < 0.5$ for $q > 0$. We remark that for $n = 2$ and the limit of $n \to \infty$, the 3D stat-mech model is in the universality of the (2+1)D quantum Xu-Moore model [39].

Appendix E: Duality between random plaquette Ising model and multi-flavor Ising model in 3D

In this section, we explicitly derive the duality between the 3D multi-flavor Ising model in Sec. IV and the 3D random plaquette Ising model (RPIM) derived for error correction in the toric code with Pauli-X/Z errors under repeated erroneous syndrome measurements [3].

The 3D RPIM derived in Ref. [3] is of the form

$$H_{\mathrm{RPIM}} = -\sum_{\mathsf{p} \in \square_{xy}} \tilde{J}_q \eta_{\mathsf{p}} u_{\mathsf{p}} - \sum_{\mathsf{p} \in \square_{xt} \cup \square_{yt}} \tilde{J}_p \zeta_{\mathsf{p}} u_{\mathsf{p}}, \tag{E1}$$

where $\tilde{J}_p = \log\sqrt{(1-p)/p}$, $\tilde{J}_q = \log\sqrt{(1-q)/q}$, $u_{\mathsf{p}} = \prod_{\mathbf{r} \in \partial \mathsf{p}} \tau_{\mathbf{r}}$, and $\eta_{\mathsf{p}}, \zeta_{\mathsf{p}} = \pm 1$ are random variables that take -1 with probability p and q, respectively. Importantly, the Hamiltonian has a constraint for $\mathsf{p} \in \square_{xy}$ on the bottom layer; the product of four spins around a plaquette is forced to be $+1$. In the language of $\mathbb{Z}_2$ gauge theory, the flux line does not penetrate the bottom boundary. The spins on the top boundary are not subject to hard constraints, and the flux line can terminate at the top boundary. The RPIM model has periodic boundary conditions in the x and y directions. Here, for the toric code on a 2D square lattice of size $L \times L$ under T rounds of repeated syndrome measurements, the RPIM model is defined on a 3D slab of size $L \times L \times T$ and consists of spins on the edges ($3L^2T$ spins in total).

Before deriving the duality, we first rewrite the partition function of the 3D RPIM in terms of the flux lines in the 3D classical $\mathbb{Z}_2$ gauge theory. Any spin configuration can be obtained by flipping a subset of spins to the -1 state from the configuration of every spin in the $+1$ state. The spin-flip creates the violation of plaquette variables along a closed loop $\mathcal{C}$ on the dual lattice, i.e. the π-flux line. This allows writing $\mathcal{Z}_{\mathrm{RPIM}}$ as

$$\mathcal{Z}_{\mathrm{RPIM}} = e^{\tilde{J}_p N_t(1-2p) + \tilde{J}_q N_{xy}(1-2q)}$$
$$\sum_{\mathcal{C}} \prod_{\mathsf{p} \in \mathcal{C}_{xy}} e^{-2\tilde{J}_p \eta_{\mathsf{p}}} \prod_{\mathsf{p} \in \mathcal{C}_t} e^{-2\tilde{J}_q \eta_{\mathsf{p}}}, \tag{E2}$$

where N_{xy} (N_t) is number of plaquettes in the xy plane (xt and yt plane), and $\mathcal{C}_{xy}$ ($\mathcal{C}_t$) is the collection of plaquettes in the xy plane (xt and yt plane) on the loop $\mathcal{C}$.

The RPIM partition function is dual to a nonlocal correlation function in a 3D Ising model with spins defined at the center of cubes in the original lattice,

$$\mathcal{Z}_{\mathrm{Ising}} = \sum_{\{s_i\}} e^{J_p \sum_{\langle i,j \rangle_{xy}} s_i s_j + J_q \sum_{\langle i,j \rangle_t} s_i s_j} \prod_{\eta_{\mathsf{p}} = -1} s_i s_j$$
$$= (\cosh J_p)^{N_t(1-p)}(\sinh J_p)^{N_t p}$$
$$(\cosh J_q)^{N_{xy}(1-q)}(\sinh J_q)^{N_{xy}q}$$
$$2^N \sum_{\mathcal{C}} \prod_{\mathsf{p} \in \mathcal{C}_{xy}} (\tanh J_p)^{\eta_{\mathsf{p}}} \prod_{\mathsf{p} \in \mathcal{C}_t} (\tanh J_q)^{\eta_{\mathsf{p}}}, \tag{E3}$$

where the link $\langle i,j \rangle$ is dual to the plaquette p, the subscript xy (t) denotes the plaquette in the xy plane (xt and yt plane), and N is the total number of spins. The two partition functions are identical up to a nonsingular prefactor provided that $\tanh J_p = e^{-2\tilde{J}_p}$ and $\tanh J_q = e^{-2\tilde{J}_q}$. Such a duality is obtained in Ref. [47].

Lastly, we show that the average $(n-1)$-th moment of the RPIM partition function, and equivalently that of $\mathcal{Z}_{\mathrm{Ising}}$, maps to the partition function of the $(n-1)$-flavor Ising model derived in Sec. IV. Here, we obtain the effective Hamiltonian by averaging over random plaquette

variables η_{p} and ζ_{p}

$$\overline{\mathcal{Z}_{\text{Ising}}^{n-1}} = \sum_{\{s_i\}} e^{\sum_{\mathsf{a}=1}^{n-1} J_p \sum_{\langle i,j \rangle_{xy}} s_i s_j + J_q \sum_{\langle i,j \rangle_t} s_i s_j}$$

$$\prod_{\langle i,j \rangle_{xy}} \left(1 - p + p \prod_{\mathsf{a}=1}^{n-1} s_i^{\mathsf{a}} s_j^{\mathsf{a}} \right)$$

$$\prod_{\langle i,j \rangle_t} \left(1 - q + q \prod_{\mathsf{a}=1}^{n-1} s_i^{\mathsf{a}} s_j^{\mathsf{a}} \right)$$

$$= \sum_{\{s_i\}} e^{-H_Z^{(n)}}. \tag{E4}$$

The effective Hamiltonian is exactly the one obtained in Eq. (D8).

We here focus on the duality between the two models in the bulk. In Appendix F, we comment on the boundary conditions in the RPIM and its dual.

Appendix F: Stat-mech model in the limit of perfect measurements

In this section, we consider the limit of perfect measurement in the multi-flavor Ising model obtained in the dual picture. We show that perfect syndrome measurements lead to a stat-mech model with open boundary conditions at the top of the slab. The resulting stat-mech model exhibits a finite-temperature ordering transition giving rise to a finite error threshold regardless of how many rounds of syndrome measurements are performed.

The couplings in the multi-flavor Ising model in the temporal direction are determined by the rate of measurement errors, $J_q = -(1/2) \log(1 - 2q)$; the coupling J_q is vanishing if $q = 0$. Thus, if the final round of measurement is perfect, the model has an open boundary condition at the top and, therefore, exhibits a transition at a finite threshold regardless of the thickness T.

The existence of a finite decoding threshold can be also understood in the original picture of the RPIM. Here, the boundary magnetic field in the dual picture translates to the boundary condition for magnetic π-flux lines. The π-flux line is allowed to terminate at the top boundary when the boundary field is nonzero. We note that the multi-flavor Ising model always has an open boundary condition at the bottom, i.e. π-flux line cannot terminate at the bottom. In this case, if the slab has a finite thickness T, there is no transition when tuning the error rate because the 2D classical $\mathbb{Z}_2$ gauge theory is always confined in the effective 2D slab. In contrast, when the fluxes are *not* allowed to terminate at both the top and the bottom, the gauge theory has a special boundary condition, which allows the transition to happen.
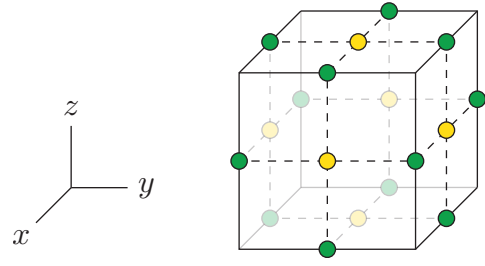

Figure 4. The unit-cell of the 3D cubic lattice Λ used to define the RBH state. The qubits are placed on the edges (green) and the faces (yellow) of Λ. The RBH state corresponds to a graph state, with the underlying graph specified by the dashed lines.

Appendix G: Toric code with repeated syndrome measurements as 3D RBH

In this section, we briefly explain how the 3D cluster state of Raussendorf, Bravyi and Harrington (RBH) [40] can be used to reason about the quantum information content of the 2D toric code under repeated noisy measurements in the presence of noise. Although the discussion in this section is focused on the 2D toric code, it readily generalizes to any CSS code, by replacing the RBH state with the corresponding foliated cluster state [61].

1. The RBH state

Consider the 3D cubic lattice Λ with qubits placed on edges and faces (Fig. 4). Let E and F denote the set of edges and faces of Λ. Let P_e (P_f) denote the Pauli operator P acting on the qubit which is placed on $e \in E$ ($f \in F$). For each edge, define $g_e = X_e \prod_{f:e \in \partial f} Z_f$, where ∂f denotes the boundary of f. Similarly, for each face $f \in F$ define $g_f = X_f \prod_{e \in \partial f} Z_e$. The RBH state is the state stabilized by g_e and g_f for all $e \in E$ and $f \in F$. Equivalently, the RBH state can viewed as the graph state [62] corresponding to the graph obtained by connecting each face qubit to the four edge qubits on the face boundary (the dashed line in Fig. 4). As such, the RBH sate $|\psi_{\text{RBH}}\rangle$ can be written as,

$$|\psi_{\text{RBH}}\rangle = U_\Lambda |+\rangle^{\otimes N} \tag{G1}$$

$$U_\Lambda = \prod_{f \in F} \prod_{e \in \partial f} \text{CZ}_{e,f} \tag{G2}$$

where $\text{CZ}_{i,j}$ stands for the controlled Z gate between qubits i and j. Thus one could prepare the RBH state by initializing every qubit in the $|+\rangle$ state and applying CZ gates along the dashed lines in Fig. 4.
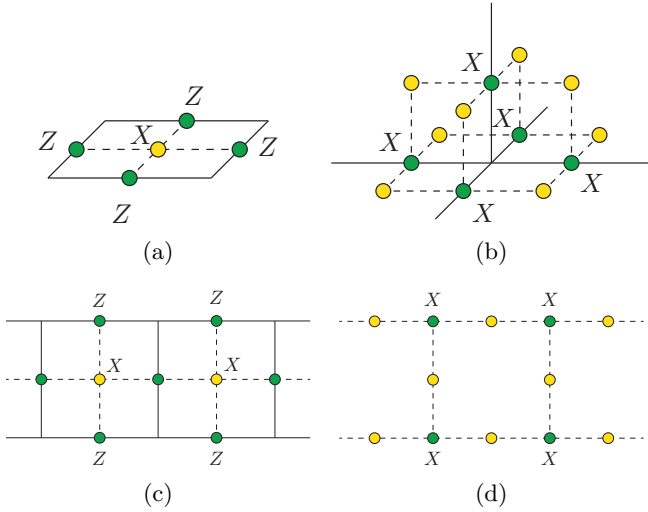
Figure 5. a) The g_f stabilizer on the boundary is equal to the toric code plaquette operator on the boundary times X_f. b) The product of four g_e stabilizers around a vertex on the boundary is equal to the toric code star operator on the boundary times X_e in the bulk. c) The product of all g_f stabilizers that lie on a vertical plane is equal to the product of the toric code Z logical operators on the top and bottom boundary times a product of some X_f in the bulk. d) The product of g_e operators where e cuts through a fixed veritcal plane is equal to the product of the toric code X logical operators on the top and bottom boundary times product of some X_f in the bulk.

2. The RBH state and the toric code

Consider the RBH state on the $L_x \times L_y \times L_z$ cubic lattice with periodic boundary conditions along the x and y axes and open boundary condition along the z axis. If one measures all qubits except the edge qubits on the top and bottom boundaries in the Pauli-X basis, the resulting state on the unmeasured qubits consists of toric code eigenstates on the boundaries, which are entangled to make a pair of logical Bell pairs [40]. To see this, note that a face stabilizer g_f on a boundary is just the toric code plaquette operator times X_f (Fig. 5a). Measuring the face qubit in the Pauli-X basis thus ensures that the unmeasured qubits are an eigenstate of the corresponding toric code plaquette operator, with the eigenvalue determined by the measurement outcome of X_f. Moreover, if one multiplies all the five edge stabilizers g_e that are stemming out of a vertex at a boundary, one obtains the star operator of the toric code on the boundary times X_e in the bulk (Fig. 5b). Thus, measuring the bulk qubits in the Pauli-X basis projects the boundary edge qubits into an eigenstate of the toric code star operator. Lastly, multiplying all the face stabilizers g_f where f lies in a fixed yz plane (or in a fixed xz plane) gives the product of the Z logical operators of the toric codes on the top and bottom boundaries times a product of X_f in the bulk (Fig. 5c). Similarly, the product of all edge stabilizers g_e

where e cuts through a fixed yz plane (or a fixed xz plane) gives the product of the X logicals times a product of X_f in the bulk (Fig. 5d). Therefore, measuring bulk qubits in the X basis projects the boundary toric codes into logical Bell pairs, with the particular Bell pairs determined by the signs of the bulk Pauli-X measurements.

3. Using the RBH state to teleport toric code eigenstates

Just like the 1D cluster state can be used to teleport a single qubit [63, 64], the 3D cluster state, i.e. the RBH state, can be used to teleport a 2D toric code state along the 3rd axis. Consider the following setup. Imagine having qubits on the edges and faces of a $L_x \times L_y \times 1$ slab of a cubic lattice Λ with periodic boundary conditions along x and y directions, and open boundaries at the $z = 0$ and $z = 1$ planes. Assume the edge qubits at the $z = 0$ plane are in a code state of the toric code $|\phi_{\text{TC}}\rangle$, and all the other qubits are initialized in the $|+\rangle$ state. First, we apply the unitary U_Λ (defined in Eq. (G2)) – which we interpret as growing a single layer RBH state – and then we measure all qubits except the edge qubits on the $z = 1$ boundary[7] in the Pauli-X basis. Doing so teleports the toric code state from the bottom boundary to the top boundary, i.e. the post-measurement state of the top edge qubits (at $z = 1$) would be $|\phi_{\text{TC}}\rangle$, up to $\pm$ sign structure differences in star and plaquette operators which can be easily determined based on the outcomes of the Pauli-X measurements. To see why, first note that one could consider an equivalent alternative scenario, where the edge qubits at $z = 0$ start in a toric code logical Bell pair state with a set of added reference qubits, instead of being initialized in $|\phi_{\text{TC}}\rangle$. Then, the same steps (i.e. applying U_Λ followed by X measurements) will result in a logical Bell pair between the reference qubits and the edge qubits at the $z = 1$ boundary. It is easy to see why this new formulation of the claim is true by viewing the original logical Bell pair between the reference qubits and $z = 0$ edge qubits as the post-measurement state of another slab of the RBH state, say between $z = -1$ (where we place the reference qubits) and $z = 0$, as described in Appendix G 2. Then the aforementioned procedure can be seen as making an RBH state on a $L_x \times L_y \times 2$ slab between $z = -1$ and $z = 1$ and then measuring all qubits in X basis except those on the top ($z = 1$) and bottom ($z = -1$) boundaries. Again, based on Appendix G 2, this would result in a logical Bell pair between the two boundaries, i.e. the reference qubits and the ones on the $z = 1$ edges. Then, teleportation follows from measuring logical operators on the reference qubits.

---

[7] Note that the edge qubits on the bottom boundary are measured as well.

4. The RBH state and repeated syndrome measurements in the absence of noise

As was explained in the previous section, the RBH state can be used to teleport the toric code along the z axis. It turns out that the procedure outlined above also measures the stabilizers of the toric code. Again, consider the setup with a $L_x \times L_y \times 1$ cubic lattice between $z = 0$ and $z = 1$ planes, where the edge qubits on the $z = 0$ are initially an eigenstate of the toric code plaquette and star stabilizers with a set of unknown eigenvalues s_i (and the rest of the qubits are in the $|+\rangle$ state). Let us first focus on the plaquette operators. Since Pauli-Z commutes with any CZ gate, the resulting state after U_Λ is applied remains an eigenstate of the plaquette operators with the same eigenvalue as before. As such, one can read off the value of plaquette stabilizers from the outcome of the Pauli-X measurements performed on the face qubits in the $z = 0$ plane (Fig. 5a). Next, we consider the star operators. Under the application of U_Λ, a star operator is dressed by the product of Pauli-Z operators on face qubits that are directly above the edge qubits participating in a star operator, since CZ XI CZ$^\dagger = XZ$. Multiplying the dressed operator by g_e, where e is the vertical edge coming out of the star, then shows that the value of the original star operator is now encoded in the five qubit Pauli-X operator shown in Fig. 5b. Hence, after measuring qubits in Pauli-X, one can read off the value of the corresponding star stabilizer from the product of the Pauli-X measurement outcomes on these five qubits. The above discussion shows that one can describe one round of syndrome measurements in a toric code state as growing a depth-one slab of RBH out of the initial state and then measuring every qubit, except those on the top edges, in the Pauli-X basis. More generally, M rounds of syndrome measurements can be described in the RBH picture by repeating the same steps M times; the first round of syndrome measurements corresponds to growing a single-layer RBH state followed by Pauli-X measurements except on the top boundary. The next round of syndrome measurements corresponds to growing another layer above the top boundary (between $z = 1$ and $z = 2$) and then measuring every qubit in Pauli-X except those on the new top boundary, and so on. The edge qubits on the $z = T$ plane then represent the state of the toric code at time $t = T$. Clearly, one can instead first grow a $L_x \times L_y \times T$ slab of the RBH state and then perform all the Pauli-X measurements throughout the slab.

5. The RBH state and repeated syndrome measurements in the presence of noise

In the last section, we explained how one can use a 3D RBH state to represent repeated syndrome measurements of the 2D toric code through time. In this section we explain how the effect of noise (including measure-
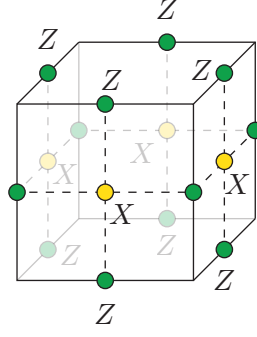

Figure 6. The product of g_f stabilizers around a cube is equal to the product of toric code plaquette operators on the top and bottom boundary times the product of four X_f in the bulk.

ment errors) can be incorporated in the RBH picture. To start, we consider the noise that can happen in one step of toric code syndrome measurements. Therefore we start with a $L_x \times L_y \times 1$ cubic lattice Λ, where the edge qubits on the $z = 0$ plane are initially an eigenstate of plaquette and star operators and rest of the qubits are initialized in the $|+\rangle$ state. As before, we first grow a single-layer RBH state state via by U_Λ, and then measure all qubits except those on the top boundary in the Pauli-X basis.

First, we explain how plaquette measurement errors can be incorporated into this picture. Since the value of a plaquette operator is inferred from the outcome of the corresponding face qubit at the $z = 0$ plane, a phase flip Pauli-Z error on that qubit before the Pauli-X measurements results in a plaquette measurement error. Importantly, such an error does not affect the actual value of the plaquette operator as it is copied from the bottom layer into the top layer after the bulk X measurements. To see this, note that for a given plaquette p, the product of g_f where f is a vertical face directly above p is equal to the product of plaquette operators on the top and bottom boundary times the product of X_f on the face qubits in between (Fig. 6). This stabilizer makes it clear why the value of the bottom plaquette operator is copied to the top layer (up to a sign depending on the X_f measurement outcomes). Since it does not involve the face qubits on horizontal planes, a phase flip on those qubits will not have any effect on the actual value of the plaquette operators on the top boundary after the bulk and bottom boundary qubits are measured in Pauli-X. As for star measurement errors, since their value is inferred from the product of X measurement outcomes shown in Fig. 5b, a phase flip Pauli-Z error on the vertical edge qubit directly above the star will result in an erroneous stabilizer measurement. For similar reasons as for the plaquette operators, such an error only flips the stabilizer measurement outcome, leaving unchanged the actual value of the stabilizer on the top boundary. Hence, this correctly represents a star operator measurement error.

To account for physical errors, it is sufficient to de-

scribe how single qubit physical Pauli-X and -Z errors can be represented in the RBH picture. To implement a physical Pauli-Z error on a qubit at time $t = 1$, we may simply apply Pauli-Z on the corresponding edge qubit on the top boundary $(z = 1)$[8]. On the other hand, a physical Pauli-X error on a qubit is represented in the RBH picture by a Pauli-Z error on the face qubit directly below the corresponding edge qubit. To see this, recall that the stabilizer shown in Fig. 6 is responsible for copying the value of a plaquette operator from the bottom boundary to the top boundary (once the bulk and bottom boundary qubits have been measured in Pauli-X). A phase flip Z_f on a vertical face qubit flips the sign of this stabilizer and thus flips the sign of the plaquette as it gets copied to the top layer. The same occurs for the adjacent plaquette, as well as the logical Z strings that pass along the corresponding edge qubit (see Fig. 5c). Therefore a phase flip Z_f on a face qubit translates into a bit flip X_e on the top edge qubit after the bulk and bottom qubits have been measured in Pauli-X.

To summarize,

- Plaquette stabilizer measurement errors correspond to Pauli-Z errors on horizontal face qubits.

- Star stabilizer measurement errors correspond to Pauli-Z errors on vertical edge qubits.

- Physical Pauli-Z errors correspond to Pauli-Z errors on horizontal edge qubits.

- Physical Pauli-X errors correspond to Pauli-Z errors on vertical face qubits.

Importantly, all types of noise in the repeated syndrome measurement setup correspond to Pauli-Z errors in the RBH picture. Since Pauli-Z commutes with CZ gates, in describing T rounds of noisy repeated syndrome measurements one may first grow a $L_x \times L_y \times T$ slab of the RBH state and then, at the end, decohere the bulk according to the noise model. This allows us to remove time completely in the RBH picture, and thus describe the (2+1)D dynamics of quantum information via a static 3D quantum system undergoing one step of decoherence.

6. Stat-mech model in the stabilizer expansion using the RBH state

In this section we sketch how one can use the RBH state to arrive at the stat-mech model describing the moments of the density matrix of the 2D toric code under repeated noisy measurement and decoherence in the stabilizer expansion formalism (Sec. IV A). The stat-mech

model in the error configuration formalism can be obtained via a similar line of reasoning. Moreover, here we focus on the stat-mech model in the bulk and leave out the details of the boundary conditions.

For simplicity, we consider the toric code under only Pauli-X errors with noisy measurement of plaquette operators, which is studied in more detail in Appendix D. Moreover, we assume the bit-flip error rate and faulty measurement rate are both equal to a given parameter p. It is straightforward to adapt the following to more general noise channels.

Let Λ^* denote the cubic lattice dual to Λ, where the vertices of Λ correspond to cubes of Λ^*, faces of Λ correspond to edges of Λ^* and cubes of Λ correspond to vertices of Λ^*. In particular, note that edge stabilizers g_e can be viewed as the face stabilizers for faces in Λ^*, which we may denote by g_{f^*}. Accordingly, the RBH state can be written as,

$$|\psi_{\text{RBH}}\rangle\langle\psi_{\text{RBH}}| = \sum_{\boldsymbol{u},\boldsymbol{v}} g^{\boldsymbol{u}} g_*^{\boldsymbol{v}}, \qquad (G3)$$

where $\boldsymbol{u}$ ($\boldsymbol{v}$) is a binary vector specifying a set of faces, i.e. membranes, in Λ (Λ^*). $g^{\boldsymbol{u}}$ and $g_*^{\boldsymbol{v}}$ are shorthand for $\prod_{f \in \Lambda} g_f^{u_f}$ and $\prod_{f^* \in \Lambda^*} g_{f^*}^{v_{f^*}}$ respectively. In general $g^{\boldsymbol{u}}$ ($g_*^{\boldsymbol{v}}$) is a Pauli string composed of Pauli-X on qubits on the surface of the membrane specified by $\boldsymbol{u}$ ($\boldsymbol{v}$) and Pauli-Z on qubits on its boundary. According to the discussion above, physical Pauli-X errors in the toric code correspond to Pauli-Z dephasing channels on vertical face qubits of the RBH state and measurement errors on plaquette stabilizers correspond to Pauli-Z dephasing channels on horizontal face qubits of the RBH state. Note that $g_*^{\boldsymbol{v}}$ does not have any Pauli-X support on the face qubits. Therefore, the corresponding decohered RBH state is,

$$\mathcal{N}(|\psi_{\text{RBH}}\rangle\langle\psi_{\text{RBH}}|) = \sum_{\boldsymbol{u},\boldsymbol{v}} e^{-\mu|\boldsymbol{u}|} g^{\boldsymbol{u}} g_*^{\boldsymbol{v}} \qquad (G4)$$

with $\mu = -\log(1 - 2p)$. Lastly, to faithfully represent the toric code state after multiple rounds of physical noise and noisy measurements, we must measure all the qubits in the bulk of the RBH state projectively in the Pauli-X basis. Note that a projective measurement channel in the Pauli-X basis is the same as a complete Pauli-X dephasing channel. Doing so will restrict the sum in Eq.(G4) to membranes $\boldsymbol{u}$ and $\boldsymbol{v}$ that do not have any boundaries in the bulk, i.e. domain walls, because otherwise $g^{\boldsymbol{u}}$ or $g_*^{\boldsymbol{v}}$ would have a Pauli-Z support and thus be annihilated under the complete Pauli-X dephasing channel. Therefore, we end up with a sum over domain walls in the 3D cubic lattice with energy penalties proportional to the domain wall sizes. This is exactly the 3D Ising model on a 3D cubic lattice, which was derived in Appendix D as well.

---

[8] Applying Pauli-Z on the corresponding edge qubit in the $z = 0$ plane corresponds to a physical Pauli-Z error at time $t = 0$.

Appendix H: Decoding transition in the repetition code

1. The repetition code

The d-dimensional repetition code consists of $N = L^d$ qubits located at the sites of a d-dimensional cubic lattice with periodic boundary conditions. The code has $I = dL^d$ check operators $g_i = Z_r Z_{r'}$, which involve qubits on the nearest neighbors r and r' to an edge i in the lattice. Since i labels the edge between these two sites, we adopt the notation $r, r' \in i$. The repetition code encodes one logical qubit of information in the GHZ state over all sites, implying that $I_0 = L^d - 1$. It follows that the check operators have $I - I_0 = (d-1)L^d + 1$ redundancies, each of which is a symmetry of the corresponding stat-mech model. The logical X operator is $\prod_r X_r$ and the logical Z operator is Z_{r_0} for any fixed site r_0.

2. Stat-mech model

The framework developed in Sec. IV A enables one to derive the entire $(n-1)$-flavor $(d+1)$-dimensional stat-mech model from just two quantities: $h_r^x(\boldsymbol{\sigma}) = \prod_i (\sigma_i)^{a_{i,r}^x}$ and $h_r^z(\boldsymbol{\sigma}) = \prod_i (\sigma_i)^{a_{i,r}^z}$. These terms couple together all the check operators with Pauli-X and Pauli-Z support, respectively, at site r. It follows that $h_r^x = 1$ and $h_r^z = \prod_{i \ni r} \sigma_i$. In other words, no check operators have Pauli-X support at any site and all check operators at edges i such that $r \in i$ have Pauli-Z support at r.

Since the repetition code only protects quantum information against Pauli-X errors, we consider the case where $p_y = p_z = 0$, which implies $J_y = J_z = 0$. Then the Hamiltonian for each flavor is

$$\mathcal{H}(\boldsymbol{\sigma}) = -J_x \sum_{t,r} \prod_{i \ni r} \sigma_{t,i} - J_q \sum_{t,i} \sigma_{t,i} \sigma_{t+1,i}, \qquad \text{(H1)}$$

where $J_q = -(1/2)\log(1 - 2q)$. The Hamiltonian $\mathcal{H}^{(n)}(\{\boldsymbol{\sigma}^{\mathsf{a}}\})$ follows from this, as defined in Eq. (43). For the 1D repetition code, the corresponding stat-mech model is a 2D Ising model. For the 2D repetition code, it is a 3D $\mathbb{Z}_2$ gauge theory.[9] Neither model orders at finite temperature when $J_q = 0$ (in which case they are reduced to a 1D Ising model and a 2D $\mathbb{Z}_2$ gauge theory respectively), indicating that they have maximum thresholds, i.e. $p_c = 0.5$, in the absence of readout noise. In the presence of readout noise, both models have sub-maximum thresholds, i.e. $p_c < 0.5$.

---

[9] In each layer, spins on edges which share a vertex are coupled, which looks like a four-body plaquette coupling on the dual lattice. Then, Eq. (H1) describes layers of 2D $\mathbb{Z}_2$ gauge theories with Ising couplings between adjacent layers. This is equivalent to a 3D $\mathbb{Z}_2$ gauge theory with all temporal links set to $+1$, which is always possible using local gauge transformations.

The symmetry group resulting from check operator redundancies is $\mathcal{R} \cong \mathbb{Z}_2^{I-I_0}$ for each flavor and $(\mathcal{R}^n \rtimes S_n)/\mathcal{R}$ for the full stat-mech model. As noted above, $I - I_0 = (d-1)L^d + 1$ for the d-dimensional repetition code. When $d = 1$, this corresponds to the global $\mathbb{Z}_2$ symmetry in each flavor of the Ising model. For $d > 1$ there are an extensive number of local symmetries generated by the spin flips of the spins along the edges of the plaquettes in the original lattice.

3. Information diagnostics

Here, we show the transition in the stat-mech model $\mathcal{Z}_n$ manifests in the information diagnostics. First, the coherent information depends on the stat-mech model we determined in Appendix H 2 as well as related partition functions with defects inserted. In particular, for $\boldsymbol{\kappa} = (\kappa_x, \kappa_z) \in \mathbb{Z}_2^2$ with κ_x and κ_z indicating the insertion of the logical X and Z operator, the Hamiltonian terms $h_{\boldsymbol{\kappa},r}^x$ and $h_{\boldsymbol{\kappa},r}^z$ change accordingly. We find that $h_{\boldsymbol{\kappa},r}^x = (-1)^{\kappa_x}$ since the logical X operator has Pauli-X support on every site, and $h_{\boldsymbol{\kappa},r}^z = (-1)^{\kappa_z} \delta_{r,r_0} \prod_{i \ni r} \sigma_i$ since the logical Z operator has Pauli-Z support only on site r_0. Here, only $h_{\boldsymbol{\kappa},r}^z$ appears in the Hamiltonian when $p_y = p_z = 0$, in which case we have

$$\mathcal{H}_{\boldsymbol{\kappa}}(\boldsymbol{\sigma}) = -J_x \sum_{t,r} (-1)^{\kappa_z} \delta_{r,r_0} \prod_{i \ni r} \sigma_{t,i} - J_q \sum_{t,r} \sigma_{t,i} \sigma_{t+1,i}. \qquad \text{(H2)}$$

The logical sector $\kappa_z = 1$ corresponds to inserting a temporal defect at site r_0 along which $\mu_x \to -\mu_x$.

The Rényi-n coherent information then takes the form

$$I_c^{(n)} = \frac{1}{n-1} \log\left(\sum_{\{\kappa_z^{\mathsf{a}}\}} e^{-\Delta F_n(\{\kappa_z^{\mathsf{a}}\})} \right), \qquad \text{(H3)}$$

where the summation over κ_z^{a} runs over 2^{n-1} choices of inserting temporal defects at site r_0 in each flavor. Here, we have performed the summation over κ_x^{a}, which cancels the constant term in Eq. (52). The Rényi coherent information can then detect the transition in the stat-mech model; $I_c^{(n)} = \log 2$ when each defect insertion costs vanishing free energy, i.e. $\Delta F_n(\{\kappa_z^{\mathsf{a}}\}) \to 0$, and $I_c^{(n)} = 0$ when each defect has a cost $\Delta F_n(\{\kappa_z^{\mathsf{a}}\}) \to \infty$. For the 1D repetition code, $\Delta F_n(\{\kappa_z^{\mathsf{a}}\}) \to 0$ in the paramagnetic phase of the corresponding 2D Ising model and $\Delta F_n(\{\kappa_z^{\mathsf{a}}\}) \to \infty$ in the ferromagnetic phase. For the 2D repetition code, $\Delta F_n(\{\kappa_z^{\mathsf{a}}\})$ is the excess free energy for inserting a magnetic π-flux line at position r_0 in the temporal direction. In the corresponding 3D $\mathbb{Z}_2$ gauge theory, the excess free energy vanishes in the deconfined phase, while it is of $O(T)$ and diverges in the confined phase.

The phase transition in the stat-mech model is also probed by the relative entropy between $\rho_{Q\mathbf{M}}$ and $\rho_{Q\mathbf{M},\boldsymbol{s}}$ where $\boldsymbol{s}$ denotes the syndromes created by applying Pauli
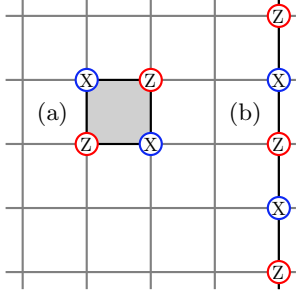
Figure 7. The XZZX code. (a) Each stabilizer is a four-body Pauli operator associated with a plaquette. Each has Pauli-X operators on the top-left and bottom-right corners and Pauli-Z operators on the top-right and the bottom-left corners. (b) The logical operators are Pauli strings along non-contractible loops involving products of Pauli-X and -Z operators.

X_r to a contiguous region in the system. For the 1D repetition code, we consider applying Pauli-X to a subregion such that $s_i = s_j = -1$ for i and j on the boundary. Then, the Rényi relative entropy is given by the two-point correlation function at the bottom of the stat-mech model, i.e. $D_s^{(n)} = -\frac{1}{n-1} \log \langle \sigma_{0,i}^1 \sigma_{0,j}^1 \rangle$. It grows with the size of the subregion $|i - j|$ in the paramagnetic phase, while it remains constant in the ferromagnetic phase. For the 2D repetition code, we consider applying Pauli-X to a subregion $\mathcal{A}$ generating $s_i = -1$ for all i along a loop $\partial \mathcal{A}$ (on the dual lattice along the boundary of the subregion $\mathcal{A}$). Then, $D_s^{(n)} = -\frac{1}{n-1} \log \langle \prod_{i \in \partial \mathcal{A}} \sigma_{0,i}^1 \rangle = -\frac{1}{n-1} \log \langle e^{i \oint_{\partial \mathcal{A}} A d\ell} \rangle$ which grows with the area of $\mathcal{A}$ in the area-law (confined) phase and with $|\partial \mathcal{A}|$ in the perimeter-law (deconfined) phase.

Appendix I: Decoding transition in XZZX code

1. The XZZX code

The XZZX code [60] involves qubits on the vertices of the square lattice, which is equivalent to a toric code up to local rotations at certain sites. It is notable for saturating the maximum error threshold of $p_c = 0.5$ when subject to only one of Pauli-X, -Y, or -Z physical errors. The XZZX code has a stabilizer group generated by the products of four Pauli operators on the vertices of a plaquette and encodes two logical qubits. Examples of stabilizer and logical operators are illustrated in Fig. 7.

Here, we consider the XZZX code on a $L \times L$ square lattice with periodic boundary conditions and discuss the stat-mech model for the code subject to Pauli-X, -Y, or -Z errors and repeated noisy syndrome measurements on all plaquettes. The set of check operators involving all plaquettes exhibits an $\mathcal{R} = \mathbb{Z}_2 \times \mathbb{Z}_2$ redundancy, giving rise to the symmetry of the stat-mech model.
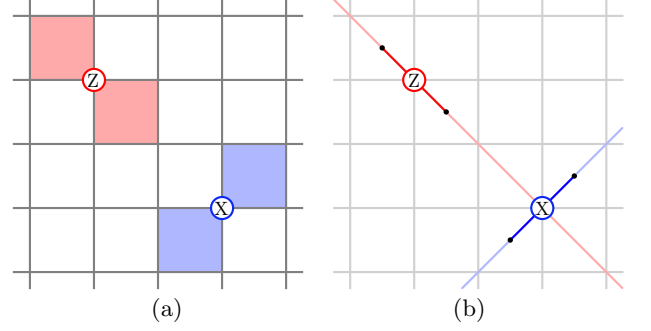

Figure 8. (a) Pauli-Z error at a site anticommutes with two diagonally-adjacent plaquette operators (colored red) that have Pauli-X support at that site. Corresponding plaquettes along the opposite diagonal have Pauli-Z support at a site, leading them to anticommute with Pauli-X errors. (b) Pauli-Z error induces two-body Ising couplings h_r^x along the diagonal from the top left to the bottom right. The analogous statement is true for Pauli-X error and h_r^z, with plaquettes coupled along the opposite diagonal.

2. Pauli-X and -Z errors

In this section, we determine $\mathcal{Z}_n$ for the XZZX code subject to Pauli-X or -Z errors. As shown in Sec. IV A, the only code-specific aspects of $\mathcal{Z}_n$ are the local spin terms h_r^x and h_r^z, so it is sufficient to compute only these terms.

As illustrated in Fig. 8(a), the upper-left and lower-right plaquettes at a site have Pauli-X support at that site, which anticommutes with Pauli-Z errors. Likewise, the upper-right and lower-left plaquettes have Pauli-Z support at that site, which anticommutes with Pauli-X errors. Taking the bottom left to be $x = y = 0$, we conclude that $h_r^x(\boldsymbol{\sigma}) = \sigma_{(x,y)} \sigma_{(x+1,y-1)}$ and $h_r^z(\boldsymbol{\sigma}) = \sigma_{(x,y)} \sigma_{(x+1,y+1)}$.

Therefore, when only Pauli-Z errors are present, each flavor of the stat-mech model consists of L uncoupled 1D Ising models running along diagonal lines from the top-left to the bottom-right. Similarly, when only Pauli-X errors are present, each flavor of the stat-mech model consists of L uncoupled 1D Ising models running along diagonal lines from the top-right to the bottom-left. These systems only order at zero temperature, leading to a decoding threshold at $p_{z,c} = 0.5$ (or $p_{x,c} = 0.5$) when there is no readout noise. Representative couplings and diagonal lines are illustrated in Fig. 8(b). In the presence of readout noise, the stat-mech model instead consists of decoupled 2D Ising models and exhibits a finite-temperature phase transition, giving rise to a sub-maximum error threshold, i.e. $p_c < 0.5$.

As derived in Sec. IV A and demonstrated in Appendix D and Appendix H, the coherent information may be formulated explicitly in terms of free energy costs of defect insertions. These defect insertions arise from the Pauli-X and Pauli-Z support of the logical operators. In
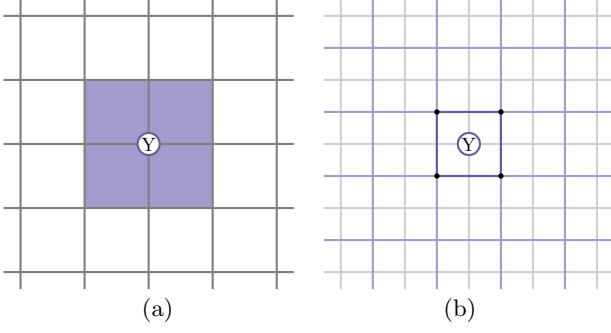
Figure 9. (a) Pauli-Y error at a site anticommutes with all four adjacent plaquettes, each of which has Pauli-X or Pauli-Z support at that site. (b) The result is a four-body plaquette Ising coupling on the lattice of classical spins.

particular, each logical operator flips the sign of the cou-

pling on one bond in each 1D Ising model.

3. Pauli-Y errors

When Pauli-Y errors are present, a four-body coupling $h_r^x h_r^z = \prod_{i \in \square} \sigma_i$ is added to the Hamiltonian, where $\square$ denotes the vertices of the plaquette dual to r. The implicated check operators and the resulting coupling are illustrated in Fig. 9(a) and Fig. 9(b), respectively.

When only Pauli-Y error is present and there is no readout noise, the decodability of the XZZX code is thus captured by a 2D plaquette Ising model, like the surface code under Pauli-Y error. This model does not order at finite temperature, giving rise to a maximum decoding threshold of $p_{y,c} = 0.5$, as expected. When we include readout error, the model is T layers of plaquette Ising models coupled by two-body Ising interactions. This model generally orders at a finite temperature, leading to a sub-maximum error threshold.