

Stabilizing Non-Abelian Topological Order Against Heralded Noise via Local Lindbladian Dynamics

Sanket Chirame^{1,*}, Abhinav Prem^{2,3}, Sarang Gopalakrishnan⁴, and Fiona J. Burnell¹

¹*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*

²*School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540, USA*

³*Physics Program, Bard College, 30 Campus Road, Annandale-on-Hudson, New York 12504, USA*

⁴*Department of Electrical and Computer Engineering, Princeton University, Princeton, New Jersey 08544, USA*



(Received 30 November 2024; accepted 12 August 2025; published 25 September 2025)

An important open question for the current generation of highly controllable quantum devices is understanding which phases can be realized as stable steady states under local quantum dynamics. In this work, we show how robust steady-state phases with both Abelian and non-Abelian mixed-state topological order can be stabilized, in two spatial dimensions, against generic “heralded” noise using active dynamics that incorporate measurement and feedback, modeled as a *fully local* Lindblad master equation. These topologically ordered steady states are two-way connected to pure topologically ordered ground states using local quantum channels, and preserve quantum information for a time that is exponentially large in the system size. Specifically, we present explicit constructions of families of local Lindbladians for both Abelian ($\mathbb{Z}_2$) and non-Abelian (D_4) topological order whose steady states host mixed-state topological order when the noise is below a threshold strength. As the noise strength is increased, these models exhibit first-order transitions to intermediate mixed-state phases where they encode robust classical memories, followed by (first-order) transitions to a trivial steady state at high noise rates. When the noise is imperfectly heralded, steady-state order disappears but our active dynamics significantly enhances the lifetime of the encoded logical information. To carry out the numerical simulations for the non-Abelian D_4 case, we introduce a generalized stabilizer tableau formalism that permits efficient simulation of the non-Abelian Lindbladian dynamics.

DOI: [10.1103/zf7y-hxtq](https://doi.org/10.1103/zf7y-hxtq)

I. INTRODUCTION

Topological phases of matter are characterized by a number of intriguing properties, including multiple locally indistinguishable ground states with subtle entanglement structure that enables them to store quantum information. However, while these phases are stable to local perturbations of the Hamiltonian that do not close the many-body gap [1–3], in $d \leq 3$, topological order (TO) does not survive at any nonzero temperature [4–12]. The fact that even a very cold environment is incompatible with low-dimensional TO raises the question of whether such orders can be stabilized at all in noisy open quantum systems. One approach to achieving this in two spatial dimensions is quantum error correction [13–22], which, when performed at regular intervals, can protect information stored

in the ground space of a topologically ordered Hamiltonian (which we henceforth refer to as a TO state) for arbitrarily long times. More specifically, in a TO state that is subjected to sufficiently weak noise for a sufficiently short time, a round of error-syndrome measurements, followed by a (typically nonlocal) recovery operation, is sufficient to fully recover the original state. Conversely, if the noise is too strong or acts for too long a time, information about the initial state is irretrievably lost in an information-theoretic phase transition. The existence of such mixed-state phase transitions has led to a surge of interest in investigating topologically nontrivial mixed states that result from local noise channels [23–47] and in defining notions of mixed-state phases of matter [48–55].

Mixed-state phases are often studied in the context of finite-time dissipative evolution; indeed, the error-correcting transition is a transition in recoverability that occurs *at finite time*. However, the most natural out of equilibrium analogs of equilibrium phases of matter, which are by definition stationary in time, are *steady states*. Clearly, one route to achieving steady-state TO is through repeated rounds of weak noise and error correction. However, because quantum error correction involves nonlocal

*Contact author: chira012@umn.edu

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

recovery processes, a dynamical regime that can be stabilized only by error correction differs sharply from a phase of matter, for which spatial locality and the Lieb-Robinson bound are crucial defining properties. To have a well-defined *steady-state phase*, we must require that the full quantum dynamics, including measurements, propagation of classical information, and the resulting feedback operations, is fully local. Against a general noise model, it is not clear that any such topologically ordered phases exist in $d \leq 3$.

The central result of this paper is that a physically motivated restriction of the noise model *does* allow for a local error-correction procedure with a nonzero threshold, and therefore to a nontrivial steady-state phase that protects quantum information. Specifically, we consider noise that is *heralded*, in the sense that one knows which sites are potentially corrupted by noise. Because heralded noise (also called erasure noise) is easier to correct than generic noise, many present-day experimental platforms rely on “erasure conversion” techniques to herald noise events; almost perfect heralding can be achieved with neutral atoms [56–62]. Remarkably, heralded noise is much easier to correct locally than generic noise. We showed previously in Ref. [40] that in one-dimensional systems with biased erasure noise, one can stabilize symmetry-protected topological (SPT) order for times that are exponentially long in system size. Here, we show that in two dimensions one can achieve much more while requiring less: for *any* heralded noise, one can construct local error-correction protocols with nonzero thresholds that stabilize both the Abelian toric code TO and a non-Abelian TO associated with the group D_4 , for times that scale exponentially in the (linear) system size. Thus, we demonstrate the existence of steady-state phases with mixed-state non-Abelian TO that are stable against general heralded noise.

The rest of this paper is structured as follows. In Sec. II, we introduce the key background concepts we use and summarize our main results. In Sec. III, we first illustrate our local error-correction protocol for the simpler case of the two-dimensional toric code, which hosts Abelian TO. We detail the heralded noise model, the correction protocol, and define the pertinent metrics we use to establish the existence of a topologically ordered steady-state phase. In Sec. IV, we provide an overview of the non-Abelian D_4 lattice model considered here (recently prepared experimentally by Ref. [63]) and discuss our local correction protocol. Here, we also introduce the stabilizer tableau method used to simulate the full Lindbladian dynamics in the non-Abelian system. Then, in Sec. V we present the results of our simulations, including the steady-state phase diagram of the D_4 model, a characterization of the transitions between distinct dynamical phases, and the effect of imperfectly heralded errors. We conclude with a discussion of open questions in Sec. VI.

II. BACKGROUND AND SUMMARY OF MAIN RESULTS

A. Background

Before turning to our central results, we briefly review the concepts of heralded noise and steady-state phases. Typical qubits in noisy intermediate-scale quantum (NISQ) devices are subjected to errors that are undetectable within the qubit subspace; in contrast, in certain settings the dominant error processes correspond to leakage errors (also called “erasure” or “flagged” errors) out of the computational subspace. These errors are then experimentally detected in real time by measurements which do not disturb the state of qubits within the logical space; qubits for which erasures are detected are then reinitialized in an arbitrary initial state, which therefore has some probability of introducing an error. Importantly, however, the location of these erasures can be stored in a classical register, giving information about the possible error locations. Systems where heralded errors are the dominant noise source include Rydberg atom arrays with biased erasure noise [58,60,64,65], trapped ions [66,67], and dual-rail superconducting cavity qubits with erasure-converted errors [59,68–71].

In the context of hardware-efficient quantum error correction, systems where erasure errors are the dominant noise sources are being actively studied given that such errors are more efficiently correctable, leading to higher thresholds and reduced overhead [56–62]. Prior work has investigated the performance of classical decoding algorithms for surface codes [72] in the presence of erasure errors; for instance, in Ref. [73] the authors found that error correction fails when the percolated clusters of erased qubits span the lattice. More recently, the authors of Ref. [74] developed a decoder for surface codes assuming perfectly heralded errors: for a given erasure pattern, all loops on the square lattice are first removed while ensuring that the erasures still span all syndromes, which are then efficiently paired up. The union-find decoder [75], an efficient and widely used decoder, extends this protocol to include the effect of unheralded Pauli errors. In either case, the protocols are nonlocal.

Here, we show that heralding can also be exploited to develop a fully *local* correction protocol, which stabilizes steady-state TO in the presence of continuous heralded noise. A key requirement for success is the ability to collapse *loops* of erasures using purely local processes, while maintaining perfect heralding of the syndromes. Our protocol does this by dynamically deforming corners of the erasure pattern in a directionally biased way. We detail in the following how this can be done in such a way that all syndromes lie along a string of flags, thereby maintaining perfect heralding and allowing the continuously acting correction protocol to stabilize a topological steady state. Locality may offer practical benefits for error correction

in terms of scalability; however, our main interest here is that such local dynamics allows us to define physically meaningful equivalence relations on the space of steady states that arise under local Markovian evolution, i.e., to define distinct robust, mixed-state steady-state phases [23,49–52].

The protocol we consider can be written as a Lindblad master equation $\partial_t \rho = \mathcal{L}(\rho)$ with two types of jump operators, corresponding respectively to the heralded noise and the recovery procedure. We have specified the recovery procedure as a “measurement-and-feedback” process; however, any jump operator can be written in this form through a polar decomposition. We are interested in the steady-state manifolds of the Lindblad superoperator $\mathcal{L}$ as a function of the noise and recovery rates. Steady states are eigenvectors of $\mathcal{L}$ with zero eigenvalue. In a stable TO phase, we expect $\mathcal{L}$ to have a nontrivial manifold of states with eigenvalues that are exponentially close in modulus to 0 (in linear system size).

We primarily define steady-state phases via the “uniformity” criterion [49]. This criterion can be stated informally as follows. Given two Lindbladians $\mathcal{L}, \mathcal{L}'$, suppose every steady state of $\mathcal{L}$ can be quickly reached by starting with a steady state of $\mathcal{L}'$ and evolving for a short time, and vice versa. Then we say that $\mathcal{L}$ and $\mathcal{L}'$ (or, equivalently, their steady-state manifolds) are in the same phase [76]. We test this criterion explicitly, starting with the steady state of the noisy evolution, turning off the noise, and checking that we rapidly approach a steady state of the noiseless evolution. Note that when this relaxation time is at most $\mathcal{O}(\text{polylog}(L))$, this criterion is equivalent to establishing two-way channel connectivity between the steady states of the noisy and noiseless evolutions, i.e., establishing that they belong to the same mixed-state phase [51].

An important caveat is in order: although we characterize phases in terms of their steady-state order (or lack thereof), in any finite-sized system, there is always a nonvanishing probability that the system ends up in an absorbing configuration (i.e., with all erasure flags occupied). Once the dynamics lands in this configuration, the system cannot escape it; hence, in any finite-sized system, strictly speaking the $t \rightarrow \infty$ steady state is always the absorbing state. However, the probability of reaching this fixed-point state can be exponentially small in the system size L [77], such that the timescale τ at which the absorbing state is reached increases exponentially with L . When we refer to a steady state in this regime, we thus always mean the steady state of the system up to these exponentially long timescales (which are beyond the range of our numerical simulations for all system sizes considered). Correspondingly, by a nontrivially ordered *steady-state phase* we mean one that does not reach the absorbing state for times that are exponentially long in the system size. On the other hand, there exist regimes in parameter space where the absorbing state is reached on a timescale (for

large enough L) that is independent of system size. We refer to such regimes, for which the absorbing state is the only possible choice of steady state, as being in the *absorbing phase*.

B. Main results

The central concept of our protocol is that it leverages information about the locations of stabilizer defects, provided by the heralded noise channel and encoded in the classical flags, to effectively confine the stabilizer defects using fully local dynamics. As a result of the competition between our local correction protocol and the heralded noise, the classical flags undergo an absorbing state transition from an active phase to the absorbing phase. The flag densities (n_X^f and n_Z^f) serve as an order parameter for such transitions [77,78]. Moreover, we show that the small values of the flag densities in the active phase precisely correspond to TO in the steady state, as directly evidenced by low logical error rates in this phase.

To illustrate our protocol in a simple setting, we first consider the ground space of the two-dimensional toric code subject to both local correction dynamics and unbiased heralded errors; the resulting steady-state phase diagram is shown in Fig. 1(a). Our local correction protocol succeeds in stabilizing a topologically nontrivial mixed steady state (the active phase in Fig. 1) for noise below a certain threshold. Specifically, we show that this mixed state (a) preserves logical information by applying the minimum-weight perfect matching (MWPM) decoder, and (b) is connected to the original pure state by a local quantum channel whose depth scales at most logarithmically with the number of qubits. Thus, the active phase of our dynamics belongs to a mixed-state phase with $\mathbb{Z}_2$ TO. We also find that a first-order phase transition separates this active phase from neighboring dynamical phases, of which two [the X -active and Z -active phases in Fig. 1(a)] preserve only classical information [79].

We then turn to our main focus, which is to show that local correction dynamics succeeds in stabilizing a mixed steady-state phase with D_4 TO for perfectly heralded noise (below a certain threshold). D_4 TO belongs to the general class of TOs which harbor pointlike *non-Abelian* anyons: whereas Abelian anyons only acquire a universal phase when braided around each other, the internal degeneracy of non-Abelian anyons can be used to construct a richer set of quantum gates, which can be exploited for topological quantum computation [72,80–82]. While states with arbitrary non-Abelian TO are challenging to prepare, those derived from solvable groups [83] permit efficient preparation via adaptive quantum circuits [84–87] and have been realized on NISQ devices [63,88–91]. In the specific case of D_4 TO, non-Abelian braiding was recently demonstrated on a trapped-ion processor [63]. Motivated in part by these experiments, we focus on TO in $d = 2$ spatial

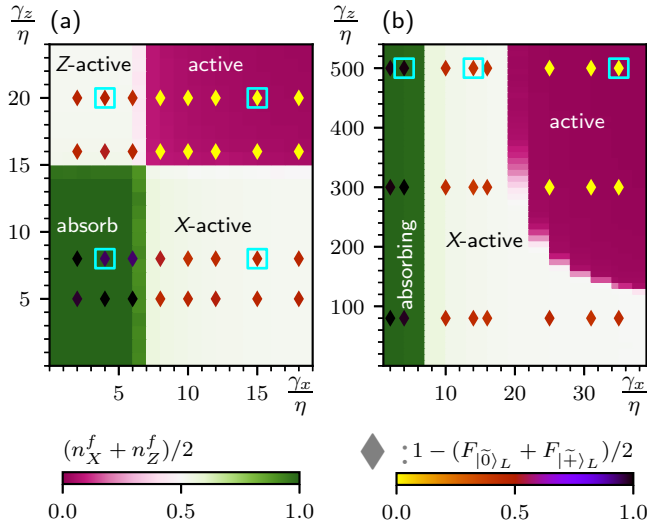


FIG. 1. Steady-state phase diagrams for (a) the toric code on a 48×48 honeycomb lattice and (b) the non-Abelian D_4 model defined on three interpenetrating 48×48 sized honeycomb lattices. The active dynamics removes the X and Z flags at rates γ_x and γ_z , respectively. The heralded depolarizing noise rate is set to $\eta = 1$. The background color indicates the combined density $\frac{1}{2}(n_X^f + n_Z^f)$ of flags evaluated at a late time ($t_f = 100$ for the toric code and $t_f = 20$ for the D_4 model) at which the densities (for parameters away from the phase boundaries) have saturated to time-independent values (see Appendix A). The logical error rates are computed for a select set of parameters depicted using the diamond-shaped symbols. These are colored according to $1 - \frac{1}{2}(F_{|0>_L} + F_{|+>_L})$, which quantifies the steady-state logical error rate averaged over two distinct initial logical states. Each data point is obtained by averaging over 100 independent Monte Carlo realizations. For both models, the active phase of the flags ($n_X^f, n_Z^f \approx 0$) leads to a topologically ordered steady-state phase, where both logical fidelities F are close to 1. This phase protects the initially encoded quantum information up to exponentially long times. In contrast, the partially active phases (only one type of flag has small density) can only protect classical information for long times, signaled by only one of the $F \approx 1$. The detailed time evolution for representative points in each of these phases, highlighted using cyan colored squares, is shown in Figs. 5 and 11.

dimensions in this paper. While we expect in principle that our protocol extends to $d = 3$ spatial dimensions *mutatis mutandis* (since it is the pointlike anyons that destabilize the phase, just as in $d = 2$), we do not pursue this here since the numerical simulations are far more challenging.

Figure 1(b) shows the resulting phase diagram, which contains a fully active phase, an absorbing phase, and a partially active phase that preserves only a classical memory [37,38,53]. As for the toric code, these are separated via first-order transitions. We provide strong evidence that the active phase of our local autonomous dynamics is in a mixed-state phase with D_4 TO: we show both that the MWPM decoder faithfully recovers quantum information

and verify numerically that the steady state is connected to a pure D_4 ground state via a local quantum channel whose depth scales at most logarithmically with the number of system qubits. We will also show that while the steady-state TO is unstable against imperfectly heralded errors, when most errors are heralded our protocol leads to a significant enhancement in the lifetime of the encoded logical information.

At first blush, it is rather surprising that local error correction succeeds for a non-Abelian anyon theory: the string operators that are required to pair-annihilate non-Abelian anyons are unitary operators whose depth is lower bounded by the linear separation between the two anyons [92]. Naïvely, this suggests that local error correction must fail. Our approach circumvents this by exploiting the fact that in anyon theories given by the quantum double $\mathcal{D}(G)$ [72,93], where G is a class-two nilpotent group (e.g., D_4), non-Abelian anyons of the same type fuse to Abelian anyons. Due to this nilpotency (or acyclicity [94]), non-Abelian anyons can be moved via depth-one unitary circuits at the expense of leaving behind a trail of Abelian anyons, which are pair-created at the end points of Pauli strings. In correcting for potential non-Abelian errors, our correction protocol allows partial information about where such strings are created, thereby enabling the confinement of both Abelian and non-Abelian errors in the long-time steady state at noise rates below a certain threshold.

A technical contribution of this work is a generalization of the stabilizer tableau [95,96] to the case of “quasistabilizers” (defined in the following), which permits efficient simulation of Lindblad dynamics for a certain class of non-Pauli stabilizers.

Finally, we comment on the values of correction rates required for these protocols and their feasibility on current erasure qubit platforms. We imagine a setup where each qubit is acted on by m gates per unit time for purposes of some desired quantum information processing task. We assume that each of these gates fails with probability p_g , resulting in a heralded error on the qubit. For small values of p_g , this can be modeled as an error process with rate $\eta_g = mp_g$. At the same time, the qubit is also affected by entangling gates that are applied during the measurement-feedback routine of our local correction protocol. The value of γ shown in Fig. 1 indicates the rate at which the protocol must check for flags; however, the measurement and feedback gates are performed only when the classical erasure flags are in an appropriate correctable configuration. This rate is given by $\alpha\gamma$ ($\alpha < 1$), with $\alpha \approx 0.03$ in the active phase of the D_4 model (see Fig. 17). If the measurement circuit fails with probability p_e , then the measurements during the local correction routine themselves contribute to the noise, with a rate $\eta_e = \alpha\gamma p_e$ [97]. Finally, qubits remaining in the idling state can decay from the computational subspace, leading to idling noise rate η_{idle} . Combining these contributions, we expect

the total noise rate to be given by

$$\eta = \eta_{\text{idle}} + mp_g + \alpha\gamma p_c. \quad (1)$$

As shown in Fig. 1(b), D_4 TO can be stabilized provided that $\gamma/\eta = R \approx 200$. Since the rates must be non-negative, Eq. (1) can be satisfied only if $p_c < 1/R\alpha$. For Yb-171 atoms [64], erasure qubits were recently demonstrated with $p_g \sim 0.02$ for two-qubit gates; taking $R\alpha \approx 10$ allows values of p_c up to 0.1, meaning that measurement-feedback must be achieved with less than 5 two-qubit gates in order to access the correction regime. In these systems $\eta_{\text{idle}} \sim 0.3 \text{ s}^{-1}$, while gate times are of the order of microseconds, suggesting that, provided the measurement error is sufficiently small, our protocol can easily control the idling errors.

III. WARM-UP: SELF-CORRECTING TWO-DIMENSIONAL TORIC CODE WITH HERALDED NOISE

Before discussing our correction protocol for the non-Abelian D_4 model, we begin by illustrating how such a protocol works for the relatively simpler case of the celebrated two-dimensional toric code [72], which hosts (Abelian) $\mathbb{Z}_2$ TO and has been experimentally realized on various NISQ platforms [98–101].

We consider a honeycomb lattice defined on a two-dimensional torus, with a qubit (together with an X and a Z flag) placed on each edge. The periodic boundary conditions (PBCs) on the $L_x \times L_y$ sized torus are implemented by making the following identifications for the position vector $\vec{r}$:

$$\vec{r} \sim \vec{r} + \frac{n_x}{3}L_x(\vec{a} + \vec{b}) + \frac{n_y}{3}L_y(\vec{a} + \vec{c}). \quad (2)$$

for $n_x, n_y \in \mathbb{Z}$ and where $\vec{a}$, $\vec{b}$, and $\vec{c}$ are the basis vectors of the honeycomb lattice, as shown in Fig. 2; these are related by $\vec{a} = \vec{b} + \vec{c}$. In the following, we set $L_x = L_y = L$, which results in a lattice with L^2 edges, $L^2/3$ plaquettes, and $2L^2/3$ vertices. Note that the toric code can be defined on any two-dimensional lattice; we choose the honeycomb lattice here to make contact with the D_4 model discussed later.

The toric code ground states $|\psi\rangle$ are those states that are simultaneously stabilized by a set of vertex B_v and plaquette A_p stabilizers, i.e., they satisfy $B_v|\psi\rangle = A_p|\psi\rangle = |\psi\rangle$ for all v, p , with

$$B_v = \prod_{j \in *v} Z_j, \quad A_p = \prod_{k \in \partial p} X_k. \quad (3)$$

Here $*v$ contains the three edges attached to vertex v , and ∂p consists of the six edges surrounding plaquette p . On the 2-torus, there are four locally indistinguishable

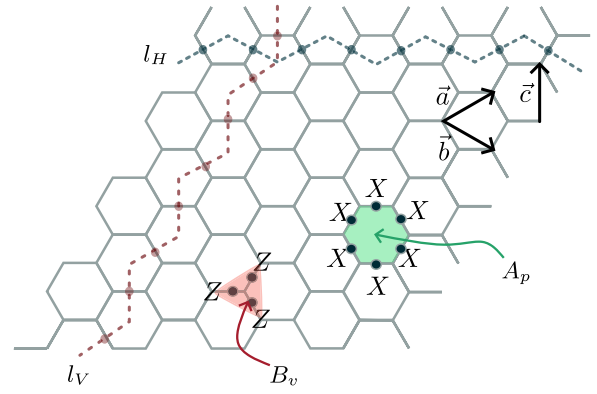


FIG. 2. Two-dimensional toric code on the honeycomb lattice: qubits live on the edges of a honeycomb lattice. A patch of 12×12 size lattice is shown, with PBCs in vertical and horizontal directions implemented according to Eq. (2). Edges involved in the vertex stabilizer B_v and plaquette stabilizer A_p are shown in red and green, respectively. The operators l_H and l_V run along noncontractible loops on the dual triangular lattice, defining the logical- $\mathcal{Z}_H$ and logical- $\mathcal{Z}_V$ operators, respectively.

ground states which can be characterized by the eigenvalues of the closed Wilson loop operators $\mathcal{Z}_H \equiv \prod_{j \in l_H} Z_j = \pm 1$, $\mathcal{Z}_V \equiv \prod_{j \in l_V} Z_j = \pm 1$, where l_H (l_V) is a horizontal (vertical) noncontractible loop on the dual lattice (see Fig. 2).

A. Noise model for unbiased heralded errors

We first introduce the noise model that we use throughout this work, which consists of unbiased, heralded errors. To model a system with heralded errors, in addition to the qubits themselves, we also introduce a classical degree of freedom, which we refer to as an erasure “flag.” These flags are represented by a pair of occupation numbers $n_{X_j}^f$ and $n_{Z_j}^f$ which are associated with each qubit j and can independently be in either occupied ($n_X^f = 1, n_Z^f = 1$) or unoccupied ($n_X^f = 0, n_Z^f = 0$) state. Here, we set both X and Z flags on all qubits to be unoccupied $n_{X_j}^f = n_{Z_j}^f = 0$ for all j in the initial state.

Here we use an error model that describes unbiased heralded noise, for which (after Pauli twirling [102]) an erasure at site j acts on the qubits via the local quantum channel:

$$\rho \rightarrow \mathcal{E}_j(\rho) = \frac{1}{4}\rho + \frac{1}{4}X_j\rho X_j + \frac{1}{4}Y_j\rho Y_j + \frac{1}{4}Z_j\rho Z_j. \quad (4)$$

While erasures typically lead to a more restricted set of errors in experiments, here we show that two-dimensional TOs can be dynamically stabilized without requiring any bias in the noise channel: this is in contrast to one-dimensional SPTs, where the strong symmetry constraint necessitated biased erasure noise [40]. Indeed, as we will

shortly demonstrate, for the two-dimensional toric code our protocol works comparably well for biased and unbiased noise. For the D_4 case studied later, we show that unbiased heralded noise represents a worst-case noise scenario for our decoder; we hence expect that more realistic noise models will lead to a significantly lower decoding overhead.

Because the errors are heralded, our noise channel also acts on the flags: upon detection of an erasure at qubit j , both X and Z flags at this site are raised by setting $n_{X_j}^f = 1$ and $n_{Z_j}^f = 1$. We model the action of this noise channel using continuous time Lindbladian evolution [103,104], given by

$$\mathcal{L}_\eta[\rho] = \sum_{\alpha, \mathbf{q}_x, \mathbf{q}_z} L_{\alpha,j}^{(\mathbf{q}_x, \mathbf{q}_z)} \rho L_{\alpha,j}^{\dagger(\mathbf{q}_x, \mathbf{q}_z)} - \frac{1}{2} \{L_{\alpha,j}^{\dagger(\mathbf{q}_x, \mathbf{q}_z)} L_{\alpha,j}^{(\mathbf{q}_x, \mathbf{q}_z)}, \rho\}, \quad (5)$$

where the Lindblad jump operators associated with the noise channel are

$$L_{\alpha,j}^{(\mathbf{q}_x, \mathbf{q}_z)} = \sqrt{\frac{\eta}{4}} f_{X_j}^+ (f_{X_j}^-)^{\mathbf{q}_x} f_{Z_j}^+ (f_{Z_j}^-)^{\mathbf{q}_z} \hat{\sigma}_j^\alpha. \quad (6)$$

Here, j labels the qubit index and $\mathbf{q}_x, \mathbf{q}_z \in \{0, 1\}$. The operators $f_{X_j}^+ (f_{X_j}^-)$ raise $n_{X_j}^f$ from 0 to 1 (lower $n_{X_j}^f$ from 1 to 0). The relation $f_{X_j}^+ f_{X_j}^- = n_{X_j}^f$ implies that the jump operator labeled by $\mathbf{q}_x = 1 (\mathbf{q}_x = 0)$ acts on the site if an X flag was already present (absent). The operators $f_{Z_j}^\pm$ act analogously on the Z flags. Here

$$\hat{\sigma}_j^0 \equiv \mathbf{1}, \quad \hat{\sigma}_j^1 \equiv X_j, \quad \hat{\sigma}_j^2 \equiv Y_j, \quad \hat{\sigma}_j^3 \equiv Z_j \quad (7)$$

are standard Pauli operators acting on the qubit at site j . The $\alpha = 0$ terms in the noise channel represent processes in which the qubit is miraculously initialized in the correct state; in this case the only effect of the erasure is to raise any flags that were not already raised at the relevant site. The remaining terms represent processes in which both flags are raised (if not already present) and a Pauli error occurs.

B. Correction dynamics for the toric code

A key feature of the toric code stabilizers in Eq. (3) is that vertex (plaquette) stabilizer defects are necessarily created in pairs, and are spatially separated via a string of contiguous Pauli- X (Pauli- Z) errors. This is a robust feature of the model because the stabilizer defects correspond to emergent anyons, which cannot be created locally. For perfectly heralded errors, this means that pairs of stabilizer defects are necessarily connected by strings of flags. Our correction protocol uses the information provided by the positions of these flags to appropriately move

stabilizer defects along such strings, thereby annihilating each stabilizer defect with its partner.

1. X flags and B_v defects

To see how the correction protocol works in practice, let us consider a vertex stabilizer defect $B_v = -1$. If only one of the three edges (say edge j) surrounding this defect has an X flag, then we can be certain that the defect is the result of an X -type error that occurred at this particular qubit. Consequently, we move the defect by applying an X operator to this qubit and simultaneously lower the X flag, as illustrated in Fig. 3(a). Similarly, at a vertex with

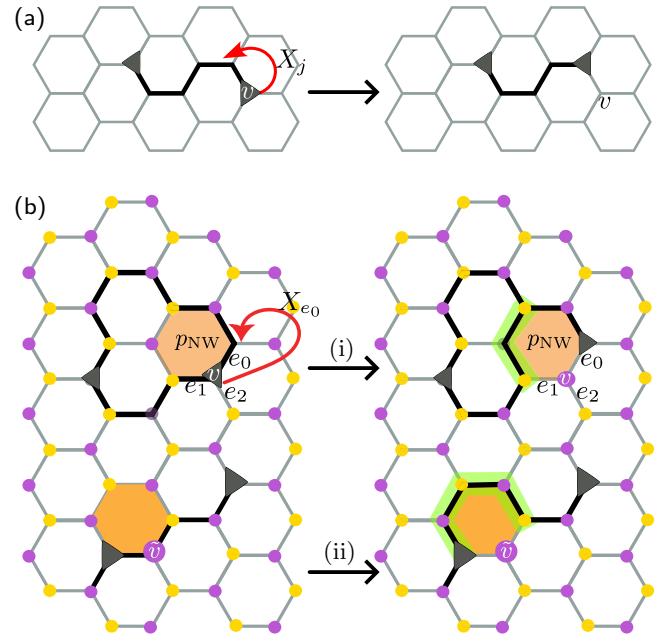


FIG. 3. X correction protocol for toric code: the edges that have X flags on them are shown using bold lines and the vertex defects are depicted by black triangles. The initial configurations on the left-hand side are updated to obtain the configuration on the right-hand side. We illustrate processes in which stabilizer defects are present near the corrected vertex; the processes with no stabilizer defects act identically on the flags, but do not act on the qubits. (a) X -leaf move: vertex v has a single edge with an X flag on it and hosts a stabilizer defect. The defect is displaced by applying X_j on the flagged edge j , after which the X flag is removed. (b) X -loop move. (i) The two edges (e_0 and e_1) emanating from a vertex v of the purple sublattice and bordering its northwest plaquette p_{NW} have X flags on them. The correction step in Eq. (9) removes these flags, raises new X flags on the remaining edges of p_{NW} (if they were not present to begin with; highlighted in green color), and displaces the vertex defect by applying X_{e_0} . (ii) The X -loop move can also be applied at vertex $\tilde{v}$ which locally hosts a flag configuration $n_{X_{e_0}}^f = n_{X_{e_1}}^f = 1$, and $n_{X_{e_2}}^f = 0$ even though the flags do not form a closed loop. As a result of this operation, the vertex defects will be paired along the alternative path highlighted in green color. The loop moves are applied only at the vertices on the purple sublattice.

no stabilizer defect and where only one of the edges has an X flag, we can simply erase the flag. These moves, which we call *leaf moves*, are implemented via the local jump operators:

$$L_{v,j,s}^{\text{TC}^{X-\text{leaf}}} = \frac{\sqrt{\gamma_x}}{2} f_{X_j}^- X_j^{\frac{1-s}{2}} (1 + sB_v) \prod_{k \in *v \setminus j} (1 - n_{X_k}^f). \quad (8)$$

Here, v labels the $2L^2/3$ vertices of the honeycomb lattice and $k \in *v \setminus j$ denotes the two edges $\neq j$ attached to vertex v . The scalar $s = -1 (s = +1)$ corresponds to the case where a vertex defect is present (absent), and γ_x parametrizes the local correction rate. The dynamics generated by this jump operator autonomously implements measurement and feedback toward a defect-free state (and requires no postselection).

In one spatial dimension, where there can be no contractible closed loops of X flags, these two processes are sufficient for stabilizing a topologically nontrivial steady state, as we previously demonstrated in Ref. [40]. However, because the leaf moves cannot remove closed loops of X -type flags, in two spatial dimensions we require additional moves to stabilize a topologically ordered steady-state. To this end, we introduce the *X -loop move*, shown in Fig. 3(b), which moves the stabilizer defects along a loop and erases the corresponding loop segment in a process which is reminiscent of Toom's rule [105].

Consider the configuration of flags and defects shown in the left-hand panel of Fig. 3(b). Here, the vertex v on the purple sublattice is found with flags on the two edges e_0 and e_1 that are north or west of it, and no flag on the third edge e_2 . Our protocol detects such configurations by checking if the flag configuration is of the form $n_{X_{e_0}}^f = n_{X_{e_1}}^f = 1$ and $n_{X_{e_2}}^f = 0$. When such a configuration is detected, the protocol lowers the flags on e_0 and e_1 , and raises new X flags on the outer edges of the northwestern plaquette p_{NW} (if they are not already present, as is the case for the vertex $\tilde{v}$ in the Figure). The B_v stabilizer at the central vertex is then measured, and if $B_v = -1$, the defect is moved up and to the right by applying X_{e_0} . The overall result of this operation is to push the string or loop of $n_X^f = 1$ flags in the northwest direction. Concurrently, the vertex defects are (a) either paired-up along the path or (b) the newly raised flags provide an alternate path on the Honeycomb lattice that is equivalent to the original error string up to a product of plaquette stabilizers A_p . Since the purpose of these moves is to shrink the loop in a preferential direction, the X -loop moves are not applied to vertices on the yellow sublattice. At a vertex v on the purple sublattice, this operation is implemented by the local jump operators:

$$L_{v,s,\mathbf{m}^x}^{\text{TC}^{X-\text{loop}}} = \frac{\sqrt{\gamma_x}}{2} X_{e_0}^{(1-s)/2} \prod_{j \in (\partial p_{NW} \setminus *v)} f_{X_j}^+ (f_{X_j}^-)^{\mathbf{m}_j^x} \times (1 + sB_v) f_{X_{e_0}}^- f_{X_{e_1}}^- (1 - n_{X_{e_2}}^f). \quad (9)$$

Here, $\mathbf{m}^x = \{0, 1\}^{\otimes 4}$ is a length-four bit string where each entry corresponds to the locations of the four qubits where potentially new X flags are raised. As a result of relation $f_X^+ f_X^- = n_X^f$, the jump operators for which $\mathbf{m}_j^x = 1 (\mathbf{m}_j^x = 0)$ act nontrivially when an X flag is already present (absent). The index $s = \pm 1$ corresponds to the measurement outcome of B_v . As before, ∂p and $*v$ represent the qubits on the edges bordering the plaquette p and connected to the vertex v , respectively.

2. Z flags and A_p defects

A similar process on the dual triangular lattice, coupled to Z flags, can be used to pair-annihilate plaquette stabilizer defects. First, if only one out of six edges surrounding the plaquette p has a Z flag [refer to Fig. 4(a)], a Z -leaf move is performed using the Lindblad jump operators:

$$L_{p,j,s}^{\text{TC}^{Z-\text{leaf}}} = \frac{\sqrt{\gamma_z}}{2} f_{Z_j}^- Z_j^{\frac{1-s}{2}} (1 + sA_p) \prod_{k \in \partial p \setminus j} (1 - n_{Z_k}^f), \quad (10)$$

where γ_z sets the rate of Z correction. This leaf move effectively moves a plaquette defect toward its partner while lowering Z flags along the error string connecting the two defects. As with X flags, loop moves are also necessary to remove Z flags that form closed loops on the dual lattice. Let us consider a plaquette p that has more than one Z flag on its northwestern edges. The possible configurations can be labeled by the occupied edges $\mathcal{E} = \{(e_0, e_1), (e_1, e_2), (e_0, e_2), (e_0, e_1, e_2)\}$, as illustrated in Fig. 4(b). For each configuration, the figure describes the locations $R[\mathcal{E}] = \{(d_0), (d_1), (d_0, d_1), (d_0, d_1)\}$ of new Z flags which are consistent with this network, meaning that errors on these edges are equivalent to errors on the original flagged edges, up to a product of vertex stabilizers B_v . For a given $\varepsilon \in \mathcal{E}$ and plaquette p , the corresponding jump operator is

$$L_{p,s,\varepsilon,\mathbf{m}^z}^{\text{TC}^{Z-\text{loop}}} = \frac{\sqrt{\gamma_z}}{2} Z_{\varepsilon(0)}^{(1-s)/2} \prod_{k \in R[\varepsilon]} f_{Z_k}^+ (f_{Z_k}^-)^{\mathbf{m}_k^z} (1 + sA_p) \times \prod_{j \in \varepsilon} f_{Z_j}^- \prod_{l \in (\partial p \setminus \varepsilon)} (1 - n_{Z_l}^f). \quad (11)$$

Here, $s = \pm 1$ denotes the stabilizer measurement outcome and $\mathbf{m}^z \in \{0, 1\}^{\otimes \dim(R[\varepsilon])}$ is a bit string that accounts for all possible configurations of Z flags on the edges in $R[\varepsilon]$. Upon detection of a plaquette defect, i.e., $s = -1$, it is always moved toward the north-most neighbor which shares the edge $\varepsilon^{(0)}$ with p .

The X (Z) loop moves allow the dynamics to shrink contractible loops of flags that may span multiple plaquettes (or vertices, for Z flags). By choosing a particular direction in which to “push” loop segments, this can be accomplished using the sequence of local moves acting on a single plaquette (vertex) and the edges in its immediate

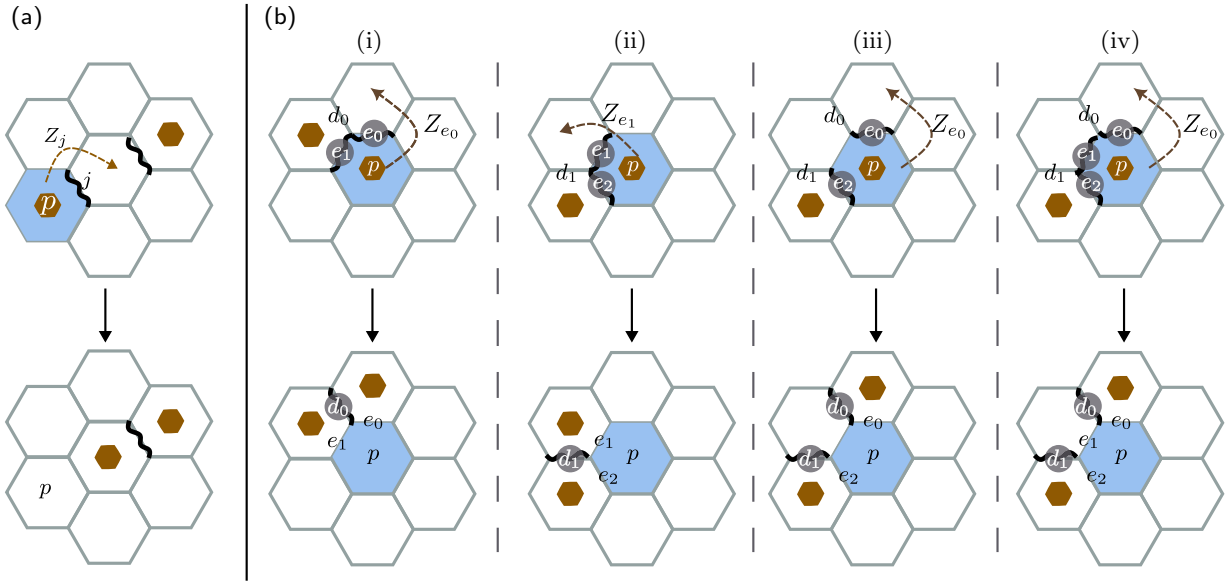


FIG. 4. Z correction protocol for toric code: the edges that have Z flags are shown using bold wavy lines. The correction operation is applied to the blue plaquette in the top row to obtain the new configuration in the bottom row. Here, we illustrate processes in which stabilizer defects are moved by acting with Z on appropriate edges (shown using dashed arrows); the processes with no stabilizer defects act identically on the flags, but do not apply the Pauli-Z operator. (a) Z-leaf move: a plaquette defect occurs on a plaquette with a single flagged edge. The correction protocol applies Z to this flagged edge, and subsequently erases the flag, moving the defect to a neighboring plaquette. (b) Z-loop moves: four candidate processes corresponding to Z flags on edges labeled by (e_0, e_1) , (e_1, e_2) , (e_0, e_2) , and (e_0, e_1, e_2) are shown here. In each configuration, new Z flags are added on edges d_0 and/or d_1 as described in the main text [refer to Eq. (11)].

vicinity. Evidently, this also means that the loop moves do not act *only* on loops; in certain configurations, they will instead rearrange the local flag and stabilizer defect configuration, as shown in Fig. 3(b)(ii). Note, however, that each of these moves only involves an $O(1)$ number of flags and stabilizer operators, thereby retaining the local nature of the dynamics.

The total superoperator implementing the correction processes is then given by

$$\begin{aligned} \mathcal{L}_C = & \gamma_x \sum_{\substack{v,s=\pm 1 \\ j \in \partial v}} \mathcal{L}_{v,j,s}^{\text{TC}^X\text{-leaf}} + \gamma_z \sum_{\substack{p,s=\pm 1 \\ j \in \partial p}} \mathcal{L}_{p,j,s}^{\text{TC}^Z\text{-leaf}} \\ & + \gamma_x \sum_{\substack{v \in \text{purple}, s=\pm 1 \\ \mathbf{m}^x \in \{0,1\}^{\otimes 4}}} \mathcal{L}_{v,s,\mathbf{m}^x}^{\text{TC}^X\text{-loop}} + \gamma_z \sum_{\substack{p,s=\pm 1, \\ \varepsilon \in \mathcal{E}(p), \mathbf{m}^z}} \mathcal{L}_{p,s,\varepsilon,\mathbf{m}^z}^{\text{TC}^Z\text{-loop}}, \end{aligned} \quad (12)$$

where the individual superoperators have the standard Lindblad form $\gamma_\Phi \mathcal{L}_\Phi[\rho] = L_\Phi \rho L_\Phi^\dagger - \frac{1}{2} \{L_\Phi^\dagger L_\Phi, \rho\}$ and the jump operators are defined in Eqs. (8)–(11). The complete dynamics is given by time evolving the density matrix under both the erasure noise Eq. (5) and the correction dynamics Eq. (12), in accordance with the Lindblad master

equation:

$$\frac{d}{dt} \rho(t) = \mathcal{L}_\eta[\rho(t)] + \mathcal{L}_C[\rho(t)]. \quad (13)$$

C. Metrics for TO and numerical results

Our goal is to use the local correction dynamics to stabilize a steady state that belongs to the same mixed-state phase as the toric code ground state. Practically, this means that the error strings must remain short, such that the stabilizer defects are effectively confined. One consequence of such confinement is that a MWPM error-correction decoder [13] will not produce logical errors (assuming that the confinement scale is small compared with the system size). In the following, we will use this decodability as an effective order parameter for the topological phase.

An alternative metric for our steady state is provided by the recent equivalence relation placed on the space of density matrices in Ref. [51]. Namely, two density matrices ρ_1 and ρ_2 belong to the same mixed-state phase if they are two-way connected via finite-depth quasilocal quantum channels, i.e., if there exist quasilocal channels Σ_{12} and Σ_{21} such that $\Sigma_{21}(\rho_1) = \rho_2$ and $\Sigma_{12}(\rho_2) = \rho_1$, where the depth of these channels is at most $\mathcal{O}(\text{polylog}(L))$. In the particular case of the two-dimensional toric code, it is known that there exists a local decoder whose threshold is quantitatively extremely close to that of the MWPM

decoder [51]; thus, here we focus only on the MWPM decoder and defer further analysis of this second metric to the D_4 active phase considered in Sec. V.

The full Hilbert space of our model, consisting of both qubits and flags, is spanned by the combined eigenstates of the A_p, B_v stabilizers and the flag occupation numbers. The noise and correction Lindbladians in Eq. (13), which generate the total dynamics, do not couple the diagonal and off-diagonal elements of the density matrix ρ in this basis. Hence, the time evolution can be simulated as a stochastic Monte Carlo dynamics of stabilizer and flag populations [78]. We initialize the system in the toric code ground state subspace, where $B_v = A_p = +1$ for all vertices v and plaquettes p , with the flags set to $n_{X_j}^f = n_{Z_j}^f = 0$ on all edges j . We then simulate the dynamics using random sequential updates of the state, where the rates of all processes are specified in the previous section. The time-evolved observables are computed by averaging over multiple independent Monte Carlo runs (see Appendix A for additional details). In what follows, we set the noise rate of the erasure channel to $\eta = 1$ and explore the steady-state phase diagram as a function of the X and Z correction rates γ_x and γ_z , respectively.

Figure 1(a) shows the numerically obtained steady-state flag density as a function of the two correction rates. As is clear from the figure, tuning the correction rates leads to four different phases: a fully *active phase* for large values of γ_x and γ_z , where the density of both types of flags saturates to a small value; an *absorbing phase* when both correction rates are small, where both X and Z flags percolate and saturate to steady-state values of $n_X^f = n_Z^f = 1$; and two *partially active phases*, where only one of the two types of flags percolates, whereas the other saturates to a low density in the steady state.

In the active phase, the stabilizer defects are corrected along the strings of erasures frequently enough such that the density of defects remains small, as shown in Fig. 5(d) at a representative point in the phase. Here, the densities of the vertex and plaquette stabilizer defects are defined, respectively, as

$$n_B^d = \frac{3}{2L^2} \sum_v \frac{(1 - \langle B_v \rangle)}{2}, \quad n_A^d = \frac{3}{L^2} \sum_p \frac{(1 - \langle A_p \rangle)}{2}. \quad (14)$$

The heralded nature of the noise and the correction dynamics also ensures that pairs of vertex (plaquette) defects are necessarily connected by strings of X - (Z)-type flags. Hence, a low density of flags necessarily corresponds to confinement of the stabilizer defects. We further illustrate this by evaluating the probability of a logical error upon pairing the defects in the time-evolved state using the MWPM decoder, ignoring the configuration of the flags. The logical state (i.e., the exact initial state within the

ground-state Hilbert space) is recovered by this decoder if the probability of both logical- X errors (logical bit-flips) and logical- Z errors (logical phase-flips) remains low. These error probabilities can be estimated by initializing the system in two different basis states in the toric code ground-state space, defined as

$$\begin{aligned} \mathcal{Z}_V |\tilde{0}\rangle_L &= \mathcal{Z}_H |\tilde{0}\rangle_L = |\tilde{0}\rangle_L, \\ \mathcal{X}_V |\tilde{+}\rangle_L &= \mathcal{X}_H |\tilde{+}\rangle_L = |\tilde{+}\rangle_L, \end{aligned} \quad (15)$$

where $\mathcal{Z}_{V/H}$ are the Wilson line operators shown in Fig. 2 (which also serve as logical Pauli- Z operators), and the logical Pauli- X operators $\mathcal{X}_{V/H}$ are equivalent line operators composed of products of X operators along closed vertical and horizontal curves on the direct lattice. The state of the qubits in a particular Monte Carlo trajectory r initialized in $|\psi_{t=0}\rangle$ can be written as

$$|\psi_r\rangle = \prod_{j \in e_x^{(r)}} X_j \prod_{k \in e_z^{(r)}} Z_k |\psi_{t=0}\rangle. \quad (16)$$

Here, $e_x^{(r)}(e_z^{(r)})$ denotes the set of edges where Pauli- X (Pauli- Z) noise is applied (up to this time) as a result of the Lindblad dynamics. The MWPM decoder generates the set of edges $d_x^{(r)}, d_z^{(r)}$ with smallest weight (counted in terms of the number of nonidentity Pauli operators) that pair-up all the defects such that the decoded state

$$|\psi_{\text{decoded}}^{(r)}\rangle := \prod_{j \in (e_x^{(r)} \oplus d_x^{(r)})} X_j \prod_{k \in (e_z^{(r)} \oplus d_z^{(r)})} Z_k |\psi_{t=0}\rangle \quad (17)$$

has no stabilizer defects. If the edges in $e_x^{(r)} \oplus d_x^{(r)}$ ($e_z^{(r)} \oplus d_z^{(r)}$) form an odd number of noncontractible loops around the torus, they act as a logical- X (logical- Z) operator on the state. Since $\{\mathcal{Z}_V, \mathcal{X}_H\} = \{\mathcal{Z}_H, \mathcal{X}_V\} = 0$, such noncontractible loops flip the eigenvalue of the logical- Z (logical- X) operator relative to the initial state, thereby taking $|\tilde{0}\rangle_L \leftrightarrow |\tilde{1}\rangle_L$ ($|\tilde{+}\rangle_L \leftrightarrow |\tilde{-}\rangle_L$). We characterize the probability of such logical flips in terms of fidelity of decoded state with respect to the initial state as

$$F_{|\psi_{t=0}\rangle} = \frac{1}{N_r} \sum_{r=1}^{N_r} |\langle \psi_{t=0} | \psi_{\text{decoded}}^{(r)} \rangle|^2, \quad (18)$$

where N_r is the number of independent Monte Carlo realizations. We numerically determine the state of logical information by pairing the stabilizer defects in the time-evolved state using the MWPM decoder implemented in the PYMATCHING-2 package [106]. The probabilities of logical- X errors ($\equiv 1 - F_{|\tilde{0}\rangle_L}$) and logical- Z errors ($\equiv 1 - F_{|\tilde{+}\rangle_L}$) are computed in terms of the fidelity of this decoded state and two distinct initial states defined in Eq. (15). In Fig. 5(d), we see that the probability of both

types of logical flips is essentially zero in the active state. The small values of logical- X (logical- Z) error rates equivalently implies the protection of initially encoded logical- Z (logical- X) information. This provides explicit evidence that quantum information is protected by the dynamical correction protocol for long times.

In contrast, the absorbing state is the unique *dead state* for the flags: when the flag density is 1, none of our correction operations can act on any sites, and the dynamics consequently no longer acts on the flags. Once the flags enter the absorbing state, they cannot leave this configuration via any of the dynamical processes. In this state, since all the correction jump operators vanish identically, the qubits evolve under the depolarizing channel and the dynamics leads to a maximally mixed steady state. This is evident in Fig. 5(a), which shows that the stabilizer defect density in this phase saturates to $1/2$. All encoded information is consequently lost, as is evident from the fact that the probability of logical flips saturates to $1 - \frac{1}{2^2} = 0.75$ since the decoder returns all possible outcomes with equal probability.

In the partially X -active (Z -active) phase, only Z (only X) flags have proliferated. Thus, our correction protocol no longer corrects Z (X) errors, but remains effective on X (Z) errors. One manifestation of this is that only

logical- Z (logical- X) information is protected in the steady state, as shown in Figs. 5(b) and 5(c). Each of these phases thus represents a *classical* memory, in which the information encoded in only one type of logical operator is dynamically protected. These phases represent intrinsically mixed topological phases which cannot be realized in the ground states of two-dimensional local gapped Hamiltonians [37,53] since they cannot be two-way channel connected to the trivial state on the 2-torus (even though they are two-way channel connected on the infinite plane).

IV. DYNAMICAL ERROR CORRECTION PROTOCOL FOR D_4 TOPOLOGICAL ORDER

We now turn to the main focus of this work: namely, exploiting heralding to achieve a steady-state phase that harbors D_4 non-Abelian TO in the presence of continuous noise processes. We focus on D_4 TO, whose anyon content is described by the quantum double $\mathcal{D}(D_4)$, for two reasons. First, this TO permits efficient preparation on a quantum processor using measurements and unitary feedback [85–87], as has been demonstrated experimentally [63]. Second, despite having 22 distinct anyons, the $\mathcal{D}(D_4)$ TO is relatively simple in the sense that D_4 is the smallest nilpotent non-Abelian group (alongside the

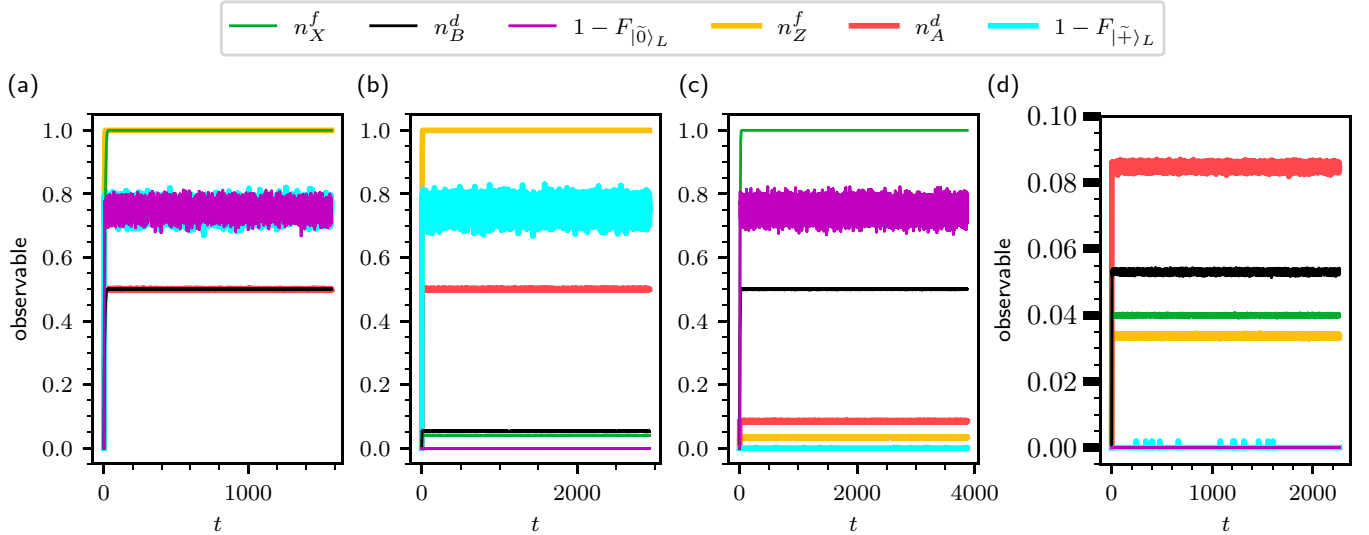


FIG. 5. Time evolution of flag and defect densities, together with the logical fidelity, for the toric code on a 48×48 honeycomb lattice for four representative points in the γ_x - γ_z phase diagram [see Fig. 1(a)]. The probabilities of logical- X errors ($\equiv 1 - F_{|0>_L}$) and logical- Z errors ($\equiv 1 - F_{|+>_L}$) are computed by decoding the state at time t using MWPM decoder [see Eqs. (15) and (18)]. The noise rate is set to $\eta = 1$, with each data point representing an average over 500 Monte Carlo runs. (a) Absorbing phase ($\gamma_x = 4, \gamma_z = 8$): where flags rapidly reach their absorbing state $n_X^f = n_Z^f = 1$, and all qubits become maximally mixed. The logical error rates $1 - F_{|0>_L}$ and $1 - F_{|+>_L}$ both saturate to their maximal value. (b) Partially X -active phase ($\gamma_x = 15, \gamma_z = 8$): the density of X flags and B defects proliferate, leading to small (large) values of logical- X (Z) error rate. (c) The analogous partially Z -active phase ($\gamma_x = 4, \gamma_z = 20$). (d) Active phase ($\gamma_x = 15, \gamma_z = 20$): here, both X and Z flag densities saturate to small values. Both logical- X (magenta colored curve) and logical- Z (cyan colored curve) error rates remain close to zero, and hence the quantum information can be recovered up to times exponentially large in system size.

quaternion group), and thus $\mathcal{D}(D_4)$ is the minimal non-Abelian TO with a structure that is amenable to local decoding. While our error-correction protocol is spiritually descended from that for the toric code, the presence of non-Abelian defects in D_4 TO necessitates significant modifications to the protocol discussed previously. The resulting phase diagram, which we discuss in detail in Sec. V, is shown in Fig. 1(b).

A. D_4 TO: quasistabilizer model and anyons

We begin with a brief overview of the properties of D_4 TO, as well as the operators that can be used to stabilize D_4 topologically ordered ground states on the lattice in a conventional qubit architecture. D_4 TO describes the phase of matter obtained by considering a discrete gauge theory with a gauge group D_4 (i.e., the dihedral group with eight elements) in $d = 2$ spatial dimensions. The pointlike excitations, or anyons, in this theory can be understood in terms of their (electric) charge and (magnetic) flux. The pure flux excitations are associated with the four nontrivial conjugacy classes of D_4 , while pure charges are associated with the four nontrivial irreducible representations (irreps) of D_4 . The remaining excitations, known as dyons, are labeled by a choice of conjugacy class together with a nontrivial representation of its centralizer. For D_4 , this leads to a total of 21 nontrivial anyons (or 22 total anyons, including the identity). A more comprehensive description of this anyon theory is provided in Appendix B.

Since D_4 is a non-Abelian group, some of these anyons are non-Abelian: when a pair of such anyons is brought together (or fused), the resulting anyon type is not uniquely determined. For example, D_4 has one two-dimensional irrep; two charges of this type can be combined in multiple ways, yielding different charge types (i.e., multiple irreps), in a manner analogous to the addition of angular momentum for irreps of $SU(2)$.

Unlike generic TOs however, the non-Abelian anyons in $\mathcal{D}(D_4)$ have fusion rules that are *acyclic* [94], meaning that when a pair of non-Abelian anyons of the same type is brought together, the possible fusion channels are all Abelian anyons. This property allows D_4 non-Abelian states to be prepared efficiently [85] and makes this system more amenable to conventional quantum error correction than generic non-Abelian TOs [20]. As we will shortly show, this property is also key to our local correction protocol for heralded noise.

In the remainder of this section, we review a specific model Hamiltonian, introduced in Ref. [107], that realizes the $\mathcal{D}(D_4)$ topological phase. This model has two advantages: first, the local Hilbert space is composed of qubits, rendering it simpler to realize on current quantum hardware [63]; and, second, the Hamiltonian can be written as

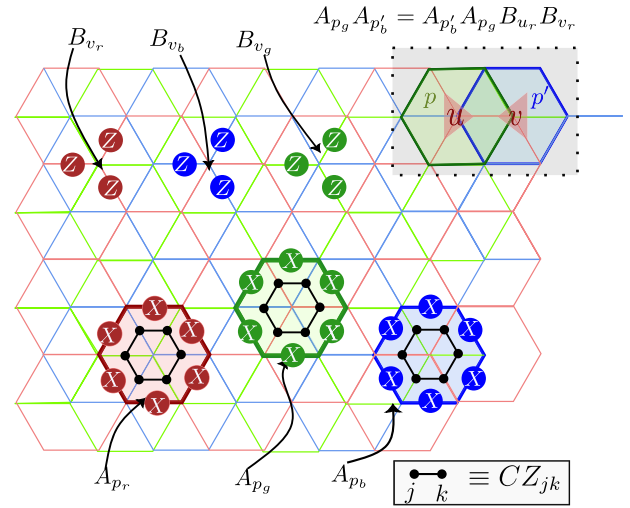


FIG. 6. Stabilizers for the D_4 model: qubits are located on edges of three interpenetrating honeycomb lattices, depicted in red, blue, and green. We use PBCs in both vertical and horizontal directions, as defined in Eq. (2), for each of these lattices. A vertex operator B_{v_c} on sublattice 1 of each lattice color is shown in the upper left; the vertex operators on sublattice 2 are products of Z operators on the remaining three edges at each vertex. A plaquette operator A_{p_c} for each color is shown on the bottom left. Pauli- X operators act on the shaded c colored edges; bold black lines indicate the two-qubit control- Z gates ($CZ_{j,k} = \frac{1}{2}(1 + Z_j + Z_k - Z_j Z_k)$) acting on pairs of qubits inside the plaquette. The box in the upper right shows plaquette operators that fail to commute, and instead satisfy Eq. (23) with $c = \text{green}$, $c' = \text{blue}$, $c'' = \text{red}$.

a sum of local “check” operators, each of which individually take eigenvalue 1 in the ground-state subspace despite not commuting in the full Hilbert space.

The model is defined on three interpenetrating honeycomb lattices with a qubit placed on each edge, as shown in Fig. 6. For convenience, the three honeycomb lattices are shown in different colors (red, green, and blue in the Figure). With each color, we associate two types of check operators that stabilize the ground states. First, the vertex check centered at vertex v_c is defined as

$$B_{v_c} = \prod_{j \in *v_c} Z_j, \quad (19)$$

where $c = r, g, b$ indicates the color of vertex v and $*v_c$ corresponds to the set of edges (of color c) incident on the vertex. For each color, these operators are identical to the vertex operators of the honeycomb lattice toric code given in Eq. (3). Second, the plaquette checks are comprised of a product of a plaquette operator of the honeycomb lattice toric code (of color c) and control- Z gates acting on the enclosed edges of the remaining two honeycomb lattices (of color c' , $c'' \neq c$). The plaquette operator acting on a red

colored plaquette p_r , for instance, is given by

$$A_{p_r} = \prod_{j \in \mathcal{I}(p_r)} CZ_{j,j+1} \prod_{k \in \partial p_r} X_k. \quad (20)$$

Here, X acts on the red edges along the boundary ∂p_r of the plaquette p_r and $\mathcal{I}(p_r)$ represents the two vertices (one blue and one green) at the center of p_r . The blue and green qubits in the interior of p_r are coupled via nearest-neighbor control- Z gates (CZ), as shown in Fig. 6. The remaining plaquette operators A_{p_b} and A_{p_g} are defined analogously.

Note that the plaquette check operators defined previously are non-Pauli “quasistabilizers.” Specifically, unlike standard stabilizer operators on an N -qubit Hilbert space, these check operators are not elements of the N -qubit Pauli group, and do not commute with each other as operators. However, they do commute within the ground-state subspace (see below for details), and stabilize the D_4 ground states. Thus, in the interest of brevity (and in an abuse of notation), we often refer to A_p and B_v operators simply as stabilizers and will emphasize the nontrivial consequences of their “quasistabilizer” nature whenever necessary.

The full Hamiltonian is given by

$$H = - \sum_{c=r,g,b} \left(\sum_{v_c} B_{v_c} + \sum_{p_c} A_{p_c} \right). \quad (21)$$

As with the toric code, it is straightforward to verify that all vertex stabilizers commute with each other and also with the plaquette stabilizers:

$$[B_{v_c}, B_{v_{c'}}] = [B_{v_c}, A_{p_{c'}}] = 0 \quad \forall p, v, v', c, c'. \quad (22)$$

However, the stabilizers on adjacent plaquettes p_c and $p_{c'}$ of different colors (see Fig. 6) do not commute; rather,

$$A_{p_c} A_{p_{c'}} = A_{p_{c'}} A_{p_c} B_{u_{c''}} B_{v_{c''}} \quad (23)$$

where $u_{c''} \in \mathcal{I}(p_c)$ and $v_{c''} \in \mathcal{I}(p_{c'})$ are vertices of the third lattice (of color $c'' \neq c \neq c'$) located at the center of plaquettes p and p' , respectively, as shown in Fig. 6. Nevertheless, Eq. (23) implies that the plaquette operators do commute when acting on states for which $B_v = 1$ at every vertex. On this subset of states, one can still find simultaneous eigenstates of all the plaquette stabilizers [87]. In particular, a ground state $|\psi_{D_4}\rangle$ of the Hamiltonian Eq. (21) satisfies

$$A_{p_c} |\psi_{D_4}\rangle = B_{v_c} |\psi_{D_4}\rangle = |\psi_{D_4}\rangle, \quad \text{for all } c, p_c, v_c. \quad (24)$$

Ground states of the Hamiltonian in Eq. (21) are thus frustration-free.

In the presence of vertex violations $B_v = -1$, on the other hand, adjacent plaquette checks fail to commute and

hence cannot be simultaneously diagonalized. In particular, in the presence of a vertex defect, it is not possible for all plaquettes to be simultaneously in their ground state. An immediate consequence is that plaquette defects can effectively be absorbed by these vertex defects: our first indication that the vertex defects are non-Abelian anyons.

Thus, unlike in the toric code, a string of Pauli- X operators on the lattice has a qualitatively different effect on the ground states $|\psi_{D_4}\rangle$ than a string of Pauli- Z operators on the dual lattice. An error Z_{j_b} on a blue edge will commute with all B_{v_c} , but anticommute with a pair of adjacent blue plaquette terms that share this edge (and similarly for other colors). Dual strings of Z -type errors thus create pairs of Abelian plaquette defects $A_p = -1$ at their endpoints; these can be pair-annihilated by reapplying the same string.

An error X_{j_b} on a blue edge will similarly anticommute with B_{v_b} at the two neighboring blue vertices, thus creating a pair of vertex defects. In addition, however, such an X error fails to commute with A_p on the red and green plaquettes that enclose this blue edge. It has the following nontrivial relation with the CZ gates present in the plaquette stabilizer:

$$X_j (CZ)_{jk} = (CZ)_{jk} X_j Z_k, \quad (25)$$

where $CZ_{jk} = \frac{1}{2}(1 + Z_j + Z_k - Z_j Z_k)$ is a control- Z gate acting on qubits j and k . This implies that the action of X_{j_b} on the ground state is

$$\begin{aligned} A_{p_r} X_{j_b} |\psi_{D_4}\rangle &= X_{j_b} Z_{k_g} Z_{k'_g} A_{p_r} |\psi_{D_4}\rangle \\ &= X_{j_b} Z_{k_g} Z_{k'_g} |\psi_{D_4}\rangle, \end{aligned} \quad (26)$$

where A_{p_r} is a red plaquette that encloses the blue edge j_b and $k_g, k'_g \in \mathcal{I}(p_r)$ denote the green edges inside this red plaquette that are connected by CZ gates to the edge j_b . Thus, in configurations with $Z_{k_g} Z_{k'_g} = +1$, measuring A_{p_r} after the X_{j_b} error acts on the ground state $|\psi_{D_4}\rangle$ will find no plaquette error, while in configurations with $Z_{k_g} Z_{k'_g} = -1$ a plaquette error will be found. Since both configurations occur in the ground state, X_{j_b} leaves this red plaquette in a superposition of having and not having a plaquette defect. Similarly, acting with a longer string of X operators on blue edges will result in a pair of vertex violations at the string’s endpoints, together with a superposition of violated and unviolated plaquette stabilizers along the string’s length, as depicted by the highlighted plaquettes in Fig. 8(a).

The underlying reason for this difference between X and Z errors stems from the fact that Z errors create Abelian (plaquette) defects, whereas the vertex defects created by X errors are non-Abelian anyons. By definition, the latter cannot be pair-created by a string of Pauli operators [63]. Instead, creating a pair of isolated non-Abelian anyons at vertices u and v of the same honeycomb lattice (say, blue)

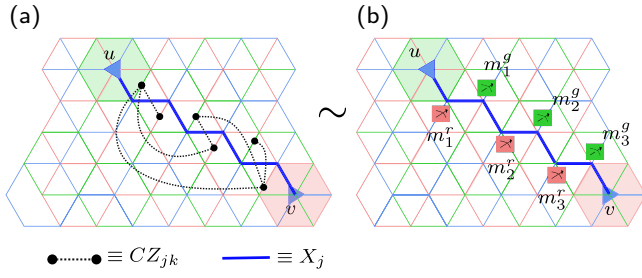


FIG. 7. Circuits for pair-creating non-Abelian vertex defects. (a) A unitary circuit $\tilde{X}_{u,v}$ as defined in Eq. (27), where X acts on bold-face blue colored edges along with a set of CZ gates (shown using dotted lines) that couple red edges with all preceding green edges for red and green edges that are connected to vertices along the path. (b) A similar effect can be achieved by first applying X on blue qubits as before, and then measuring the A_{p_g} and A_{p_r} operators indicated by green and red colored boxes, respectively. If the outcomes $m^g, m^r \equiv 1$ are postselected on the condition of observing no plaquette defects, the effect is the same as applying the non-Abelian anyon string in (a). Importantly, any measurement outcome leads to a state in which reapplying the Pauli- X string annihilates all non-Abelian defects, but leaves the system in a superposition of states with different numbers of Abelian anyons along the string's length. Both of these circuits generate $B_u = B_v = -1$ vertex defects at the endpoint of the string, with the highlighted plaquettes (centered around the endpoints) in a superposition state such that $\langle A \rangle = 0$.

requires a unitary circuit with linear depth (in distance between u and v):

$$\tilde{X}_{u,v} = \prod_{\ell_b \in C_b} X_{\ell_b} \prod_{\substack{\ell_r > \ell_g \\ \ell_r, \ell_g \in C_{rg}}} CZ_{\ell_r, \ell_g}. \quad (27)$$

Here, C_b is an open string along the blue honeycomb lattice that connects vertices u and v , and $\ell_b \in C_b$ denotes links along this path [see Fig. 7(a)]. The set C_{rg} consists of those links on the red and green lattices which are incident on vertices along C_b and are also either enclosed by or on the boundary of two plaquettes that enclose vertices in C_b [108]. The notation “ $\ell_r > \ell_g$ ” indicates those pairs of red (connected to vertices v_r along C_b) and green (connected to vertices v_g along C_b) links in C_{rg} for which v_r is to the right of v_g . We can similarly define open-string operators that pair-create isolated red and green vertex defects.

Using the definition of the plaquette stabilizers and the relation in Eq. (25), it is straightforward to verify that $\tilde{X}_{u,v}$ creates defects only at the string's endpoints: if p is a plaquette that encloses a pair of edges in C_b , then $[A_p, \tilde{X}_{u,v}] = [A_{p_g}, \tilde{X}_{u,v}] = 0$ and hence these stabilizers remain in their ground state after applying $\tilde{X}_{u,v}$. Thus, plaquette defects can occur only on the two plaquettes that enclose the vertices u and v (for which $\langle A_p \rangle = 0$ after applying the $\tilde{X}_{u,v}$ operator.)

In other words, to create only a pair of vertex defects, a Pauli- X string acting on blue edges must be decorated with a sequence of CZ gates that couple green and red edges along the string's full length; the non-Abelian anyon string is thus inherently nonlocal. A bare undecorated string of X operators, in contrast, cannot simply create a pair of pointlike excitations; rather, it creates a veritable anyon soup (i.e., a linear combination of states with all different numbers of anyons) along the string's length; the acyclic nature of the fusion rules in the $\mathcal{D}(D_4)$ anyon theory is reflected in the fact that this soup contains only Abelian anyons. Note, however, that measuring the plaquette operators away from the string's endpoints has some probability to project all of these plaquettes onto their ground states, leaving only a pair of pointlike defects at the string's endpoints. In other words, these measurements, applied after the X string, have some probability of effectively acting like the non-Abelian anyon string operator as shown in Fig. 7(b) [109]. Even for general measurement outcomes, reapplying the Pauli- X string after measurements does not return the stabilizers to their initial states. Thus, in a situation with active dynamics involving plaquette measurements, it is also not safe to assume that what began as a simple string of Pauli- X errors acting on the ground state can be undone simply by reapplying this same string. This is the essence of the complications that we will encounter, relative to the toric code, in constructing our decoder in the following.

The full spectrum of pointlike excitations (or superselection sectors) in this model can be described in terms of combinations of the three colors of vertex and plaquette defects, yielding the 22 anyons of the D_4 TO [87]. The correspondence between the lattice defects discussed above and the conventional labels for these anyons by a conjugacy class and irreducible representation of its centralizer is summarized in Appendix B (see Table I).

B. Error-correction protocol

We now describe the local correction protocol that we use to stabilize a steady state that lies in the same mixed-state phase as the ground space of the D_4 Hamiltonian Eq. (21), for noise rates below a finite threshold. Note that the heralded noise channels are the same as in the toric code [Eqs. (5) and (6)], aside from additional color indices. In the following, the color index c is suppressed for convenience wherever it does not appear explicitly.

As with the correction protocol for the two-dimensional toric code (see Sec. III B), our correction protocol for D_4 involves two types of processes. The first are leaf moves, which shrink strings of flags from their endpoints inward, while simultaneously moving any nearby stabilizer defects. The second are moves that can break open loops of flags, while pushing the corresponding stabilizer

defects in a fixed direction (chosen to be the northwestern corner here) such that they necessarily remain adjacent to at least one flagged edge. In the D_4 case, however, a new feature arises due to the fact that the plaquette operators A_p are “decorated.” More precisely, measuring plaquette stabilizers between the application and removal of Pauli- X errors can create a situation where removing the X error leaves behind a plaquette defect (Z error), as described in Eq. (26). When all X operators are flagged, this can be avoided by only measuring plaquette checks when none of the interior edges have X flags. However, X loop moves introduced in the toric code protocol can generate configurations where the flags track the locations of previous X operators only up to closed Pauli- X loops; thus, we cannot be certain that correcting X errors will not introduce new plaquette defects. This leads to two fundamental differences between our dynamical protocol for the D_4 model and its toric code counterpart. First, the dynamics of stabilizer defects created as a result of X -type and Z -type errors (and, therefore, the dynamics of the X and Z flags) are no longer independent of each other, since we must add new Z flags as we correct X errors. Second, the nontrivial commutation relation of X_j with the surrounding A_p means that the dynamics of plaquette defects is no longer diagonal in the stabilizer basis, and the conventional stabilizer tableau method fails.

1. X flags and B defects

We begin by discussing how the correction protocol acts on X flags and the associated vertex defects. The basic idea underlying the correction protocol for X flags and vertex defects is similar to that used in the toric code. At a vertex where only one of the edges hosts an X flag, we first measure B_v , then apply X on the flagged edge if $B_v = -1$, and finally remove the flag: exactly as we did for the toric code. The leaf operations must be supplemented with an X -loop move to ensure that the protocol can remove contractible loops of flags. As noted previously, however, these correction steps may inadvertently create new Z errors: thus, after this correction step, we must also raise Z flags to indicate the locations of the possible new Z errors, as shown in Fig. 8.

We now describe these two types of moves in more detail. The loop moves act on the X flags and B_v defects exactly as for the toric code. An example of the resulting loop moves, including the positions of the new Z flags, is shown in Fig. 8(b). There, a vertex v_b on sublattice 1 (sublattice formed by southeast corners of blue honeycombs) of the blue honeycomb lattice has X flags on the blue edges above it (e_0) and to its left (e_1) and no X flag on the remaining edge e_2 , ensuring that a loop move will be applied. The loop move removes the two X flags on these edges, adds X flags on the remaining four edges of the plaquette (if these are not already present), and pushes the vertex defect

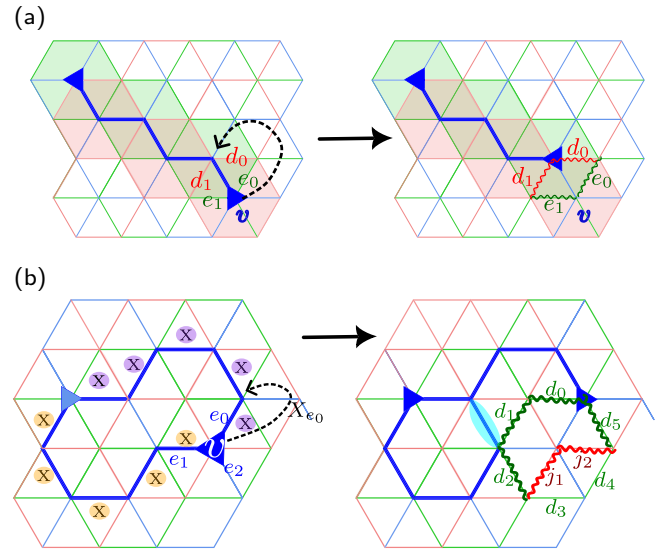


FIG. 8. Correction protocol for X errors in the D_4 model. (a) X -leaf move: a string of X erasure errors, shown using thick blue lines, has vertex defects $B_v = -1$ at each endpoint. The highlighted plaquettes of red and green color along the length of the string are in a superposition state, as described by Eq. (26). A blue vertex v_b has a single X flag on the surrounding blue edges. This flag is removed, and the vertex defect (if present) is displaced by applying X to the corresponding qubit. Following this, new Z flags are raised on edges e_0 , e_1 , d_0 , and d_1 (wavy lines) to herald potential plaquette defects created during this correction step. (b) X -loop move. Left: A configuration with X flags on both edges (e_0 and e_1) above and to the left of vertex v , which will result in an X -loop move. Here, we show a configuration that has a loop of blue X flags, along with two possible X error strings (colored in purple and yellow) consistent with the flags that can generate the pair of defects shown using blue triangles. Right: The loop move removes the flags on e_0 and e_1 , and displaces any vertex defects up and to the right by applying X_{e_0} . Wavy lines indicate the edges $\Xi[v] = \{d_0, d_1, d_2, d_5, j_1, j_2\}$ surrounding the vertex where new Z flags are raised (if not already present). In addition, as in Fig. 3, new X flags on the outer edges of the plaquette are raised if they were not already present (in this case, a single new X flag, highlighted in cyan).

upward and to the right by applying X_{e_0} . If the defect was created by an X_{e_0} error, then we have exactly undone this error. However, if the defect was generated by an X_{e_1} error, this correction leads to a net application of $X_{e_0}X_{e_1}$ to the initial state, in spite of the fact that the edges e_0 and e_1 no longer carry X flags. From Eq. (26), we see that subsequent measurements of plaquette stabilizers (which occur during the Z correction step described later) on this state will result in

$$\begin{aligned}
 (1 + s_g A_{p_g}) X_{e_0} X_{e_1} |\psi_0\rangle &= (1 + s_g Z_{j_1} Z_{j_2}) X_{e_0} X_{e_1} |\psi_0\rangle, \\
 (1 + s_r A_{p_r}) X_{e_0} X_{e_1} |\psi_0\rangle &= (1 + s_r Z_{d_0} Z_{d_5}) X_{e_0} X_{e_1} |\psi_0\rangle, \\
 (1 + s'_r A_{p'_r}) X_{e_0} X_{e_1} |\psi_0\rangle &= (1 + s'_r Z_{d_1} Z_{d_2}) X_{e_0} X_{e_1} |\psi_0\rangle,
 \end{aligned} \tag{28}$$

where $s_j \in \{1, -1\}$ denotes the measurement value obtained for A_{p_j} . Here, A_{p_g} is the green plaquette enclosing the vertex v_b and $A_{p_r}, A_{p_r'}$ are green plaquettes centered on the endpoints of the edges e_0, e_1 , respectively. The action of $I \pm ZZ$ based on the measurement outcomes s_g, s_r, s_r' can generate new defects on the adjacent plaquettes.

Instead of attempting to avoid creating such defects, our strategy is simply to ensure that they are properly heralded, allowing them to be corrected at a later stage. We therefore add Z flags to each of the six qubits in $\Xi(v_c) := \{d_0, d_1, d_2, d_5, j_1, j_2\}$, thereby heralding any Abelian defects that may potentially appear as a result of this X correction process. The jump operator implementing this sequence at a vertex on sublattice 1 of a c -colored lattice can be written as

$$L_{v_c, s, \mathbf{m}^z}^{D_4^{X-\text{loop}}} = L_{v_c, s, \mathbf{m}^z}^{\text{TC}_{(c)}^{X-\text{loop}}} \prod_{k \in \Xi(v_c)} f_{Z_k}^+ (f_{Z_k}^-)^{\mathbf{m}_k^z}, \quad (29)$$

where $L^{\text{TC}(c)}$ is the toric code jump operator defined in Eq. (9) and $\Xi(v_c)$ represents the set of six qubits in Eq. (28), where new Z flags are added if not already present. The bit string $\mathbf{m}^z \in \{0, 1\}^{\otimes 6}$ accounts for all possible configurations of preexisting Z flags on the edges in $\Xi(v_c)$ where we recall the relation $f_z^+ f_z^- = n_z^f$. Similar to the toric code case, the loop moves are applied to the vertices of only one of the sublattices of each c -colored honeycomb lattice.

The process for adding flags after a leaf move is similar. As noted before, the vertex defects are corrected either along the path they were generated or along an alternative path that is equivalent to the original one, up to a product of stabilizers. While the former correction leaves the state unchanged, the latter results in the action of closed loops of Pauli X operators. Thus, new Z flags must also be added after leaf moves. Specifically, after a leaf move that results in overall application of X_{j_b} , measuring A_{p_g} or A_{p_r} on one of the plaquettes centered on the two endpoints of the edge e_{j_b} gives

$$\begin{aligned} (1 + s_g A_{p_g}) X_{j_b} |\psi_0\rangle &= (1 + s_g Z_{d_0} Z_{d_1}) X_{j_b} |\psi_0\rangle, \\ (1 + s_r A_{p_r}) X_{j_b} |\psi_0\rangle &= (1 + s_r Z_{e_0} Z_{e_1}) X_{j_b} |\psi_0\rangle. \end{aligned} \quad (30)$$

After a leaf move, new Z flags are therefore raised on the four edges $\Omega(j) = \{e_0, e_1, d_0, d_1\}$, as shown in Fig. 8(a); these ensure that any such plaquette defects are heralded. Overall, the modified X -leaf-correction procedure is given by

$$L_{v_c, j, s, \mathbf{m}^z}^{D_4^{X-\text{leaf}}} = L_{v_c, j, s}^{\text{TC}_{(c)}^{X-\text{leaf}}} \prod_{k \in \Omega(j)} f_{Z_k}^+ (f_{Z_k}^-)^{\mathbf{m}_k^z}, \quad (31)$$

where $L^{\text{TC}(c)}$ represents the analogous toric code jump operator centered at vertex v_c defined in Eq. (8) and the

second term raises new Z flags on qubits in the set $\Omega(j)$ if they were not present before, which is accounted for by the bit string $\mathbf{m}^z \in \{0, 1\}^{\otimes 4}$ and relation $f_z^+ f_z^- = n_z^f$.

2. Z flags and A defects

Recall that a Z_{k_c} ($c = r, g, b$) error on a single edge anticommutes with the two A_{p_c} plaquette operators sharing that edge; thus, strings of Z errors create pairs of Abelian anyons that always fuse back to the vacuum $A_{p_c} = +1$ when brought together. In the previous section, we observed that new Z flags and plaquette defects emerge as a result of vertex correction processes. The flags that herald these new defects are only raised once the X flags in the vicinity of the plaquette are removed. If we instead measure the plaquette before the complete set of Z flags is available, the defects may get pushed in a direction incompatible with that of their paired partner. We avoid this by requiring that the X flags on qubits enclosed inside a plaquette p_c must be removed *before* measuring A_{p_c} . This ensures that the potential defect $A_{p_c} = -1$ will be perfectly heralded with respect to its partner defect.

This modified leaf-correction procedure at a plaquette p_c that hosts a single Z flag on one of the qubits j that bounds the plaquette is described by

$$L_{p_c, j, s}^{D_4^{Z-\text{leaf}}} = L_{p_c, j, s}^{\text{TC}_{(c)}^{Z-\text{leaf}}} \prod_{k \in *\mathcal{I}(p_c)} (1 - n_{X_k}^f), \quad (32)$$

where $L^{\text{TC}(c)}$ represents the corresponding toric code jump operator in Eq. (10) and $*\mathcal{I}(p_c)$ contains the six edges enclosed inside the plaquette p_c , as shown in Fig. 9(a); the second term restricts the operation to configurations where X flags are absent on these edges. This must be supplemented with the corresponding loop moves implemented using

$$L_{p_c, s, \varepsilon, \mathbf{m}^z}^{D_4^{Z-\text{loop}}} = L_{p_c, s, \varepsilon, \mathbf{m}^z}^{\text{TC}_{(c)}^{Z-\text{loop}}} \prod_{k \in *\mathcal{I}(p)} (1 - n_{X_k}^f). \quad (33)$$

Here, $L^{\text{TC}(c)}$ represents the toric code jump operator discussed in Eq. (11) and the second term again restricts the operation to configuration without any X flags in the interior of the plaquette p_c .

C. Simulation using non-Pauli stabilizer tableau

We now describe our simulation protocol for the complete Lindbladian dynamics, which includes the heralded noise as well as the correction operators discussed previously, and is given by

$$\frac{d}{dt} \rho = \sum_{c=r,g,b} (\mathcal{L}_{\eta_c}(\rho) + \mathcal{L}_{C_c}(\rho)), \quad (34)$$

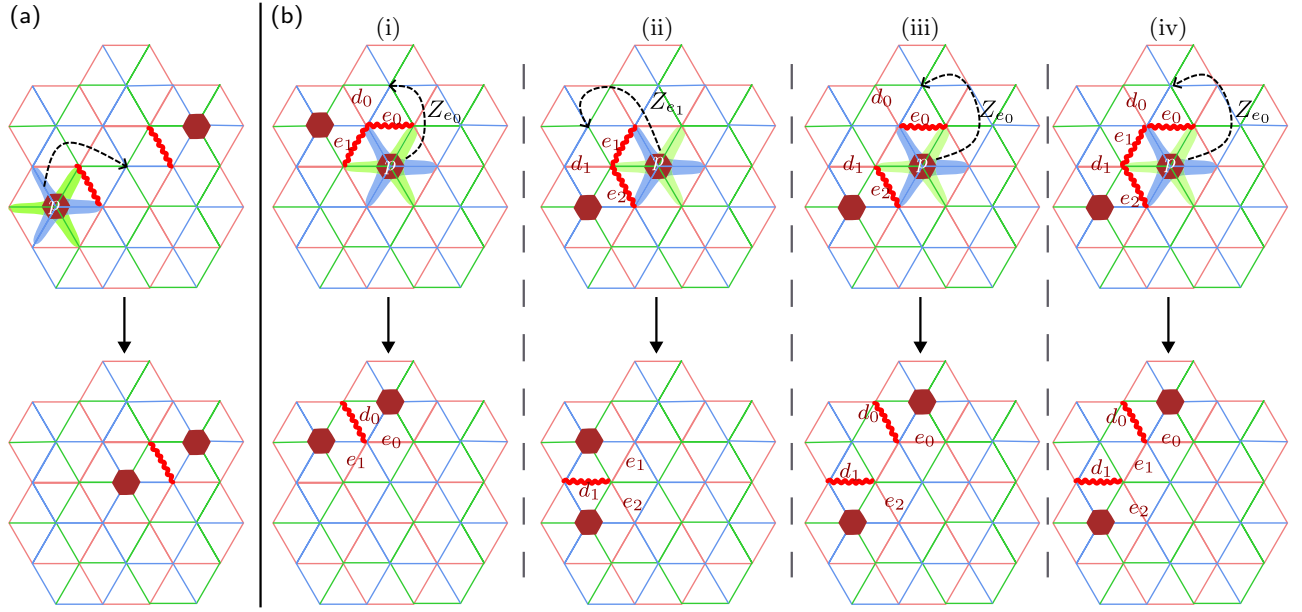


FIG. 9. Correction protocol for Z errors in the D_4 model. The dots indicate the plaquette defect $A_{p_r} = -1$, and wavy lines represent a Z flag on the edge. The correction proceeds at a plaquette p only if X flags on all qubits (highlighted in green and blue) connected to the vertices $\mathcal{I}(p)$ enclosed by the plaquette p are in the state $n_x^f = 0$. (a) Z -leaf move: the defect on plaquette p is displaced in the direction of the edge that has Z flag on it. (b) Z -loop moves: four different configurations where plaquette p has Z flags on edges (e_0, e_1) , (e_1, e_2) , (e_0, e_2) , and (e_0, e_1, e_2) are shown in subpanels (i)–(iv). The plaquette defect is pushed toward the topmost neighbor via application of Z . New Z flags (shown in red) are raised on the edges d_0, d_1 outside the plaquette, similarly to the toric code protocol (see Fig. 4).

where $\mathcal{L}_{\eta_c}$ applies heralded depolarizing noise discussed in Eq. (6) to every qubit and $\mathcal{L}_{C_c}$ describes the combined correction process

$$\begin{aligned} \mathcal{L}_{C_c} = & \gamma_x \sum_{\substack{v_c, s = \pm 1 \\ j \in \mathcal{I}^*(v_c), \\ \mathbf{m}^z \in \{0,1\}^{\otimes 4}}} \mathcal{L}_{v_c j, s, \mathbf{m}^z}^{D_4 X\text{-leaf}} + \gamma_z \sum_{\substack{p_c, s = \pm 1 \\ j \in \partial p_c}} \mathcal{L}_{p_c j, s}^{D_4 Z\text{-leaf}} \\ & + \gamma_x \sum_{\substack{v_c \in \text{sublattice-1}, \\ s = \pm 1, \\ \mathbf{m}^x \in \{0,1\}^{\otimes 4}, \\ \mathbf{m}^z \in \{0,1\}^{\otimes 6}}} \mathcal{L}_{v_c, s, \mathbf{m}^x, \mathbf{m}^z}^{D_4 X\text{-loop}} + \gamma_z \sum_{\substack{p_c, \\ s = \pm 1, \\ \varepsilon \in \mathcal{E}(p_c), \\ \mathbf{m}^z}} \mathcal{L}_{p_c, s, \varepsilon, \mathbf{m}^z}^{D_4 Z\text{-loop}}. \end{aligned} \quad (35)$$

Here individual Lindblad superoperators $\gamma_\Phi \mathcal{L}_\Phi = L_\Phi \rho L_\Phi^\dagger - \frac{1}{2} \{L_\Phi^\dagger L_\Phi, \rho\}$ are defined in terms of the jump operators detailed in Eqs. (29) and (31)–(33).

In our protocol, the flags undergo classical dynamics and can thus be simulated using standard Monte Carlo techniques (see [78]). However, as mentioned before, one complication of the correction protocol, relative to that required for the toric code, is that the dynamics is no longer diagonal in the quasistabilizer basis. This is because applying Pauli- X operators creates a superposition of different configurations of Abelian anyons. Hence, simulating the full dynamics of the quantum system requires additional techniques. Here, we adapt the stabilizer tableau method

[95,96] and describe how an appropriate generalization allows us to simulate the dynamics of the quasistabilizer defects. Specifically, this method allows us to simulate individual measurement trajectories of the quasistabilizers, which we then average over independently generated measurement trajectories using a Monte Carlo approach to compute physical observables (see Appendix A). The key difference in our approach, relative to usual stabilizer tableaux, is that the quasistabilizers defining the D_4 state are *noncommuting* operators that are not elements of the Pauli group. We outline how this complication can be treated here, with further details provided in Appendix C. As before, we slightly abuse the notation and refer to A_p and B_v as stabilizers whenever it is clear from the context.

Our system begins in a state stabilized by the A_p and B_v operators (the color index c is implicit here), satisfying

$$A_p |\psi_0\rangle = B_v |\psi_0\rangle = |\psi_0\rangle, \quad \text{for all } v, p \quad (36)$$

[see Eqs. (19) and (20)] as well as $n_{x_j}^f = n_{z_j}^f = 0$, for all edges j . The entire noise and correction dynamics involves the action of the Pauli operators X, Z , and measurements of A_p and B_v . The basic approach that we use is to track how the application of Pauli operators and measurements modifies the stabilizers over the course of the simulation. In other words, we track individual trajectories in the time evolution of the qubits by time evolving these stabilizers.

Since the flags undergo classical dynamics that remains diagonal in their population basis, this is sufficient to allow us to simulate individual trajectories of the full system.

To determine the effect of applying a unitary operator, we use the fact that if $S|\psi\rangle = |\psi\rangle$ then $USU^{-1}U|\psi\rangle = U|\psi\rangle$; thus, the stabilizer of the new state $|\psi'\rangle := U|\psi\rangle$ is given by $S' = USU^{-1}$. Applying Pauli Z_k hence gives us

$$B_v \xrightarrow{Z_k} B_v, \quad \pm A_p \xrightarrow{Z_k} \mp A_p, \quad (37)$$

where k is one of the outer edges of plaquette p . Similarly, applying Pauli X_k transforms the vertex operators as

$$\pm B_v \xrightarrow{X_k} \mp B_v, \quad (38)$$

where v represents vertices (of same color as edge k) at either endpoints of edge k .

While applying X only updates the vertex stabilizers up to a sign, it modifies the form of the plaquette stabilizers. For example, applying X_{e_1} updates the plaquette stabilizers as

$$A_p \xrightarrow{X_{e_1}} A'_p := Z_{e_0} Z_{e_2} A_p \quad (39)$$

for any plaquette p that encloses the edge e_1 , with e_0, e_2 being the adjacent edges (of a different color) that share the central vertex, as shown in Fig. 10. Here, we have used the relation $X_j CZ_{j,k} = CZ_{j,k} X_j Z_k$ and the locations of CZ gates in A_p to determine the modified plaquette stabilizer. As with the original plaquette operators, note that while the new plaquette terms do not commute as operators, by construction they commute when acting on the modified state $|\psi'\rangle = X_{e_1}|\psi\rangle$.

Next, to determine the effect of measurements, we use the fact that at any point during the simulation, the instantaneous stabilizers have the form $\{\pm B_v, \pm \prod_{k,p} Z_k A_p\}$. These

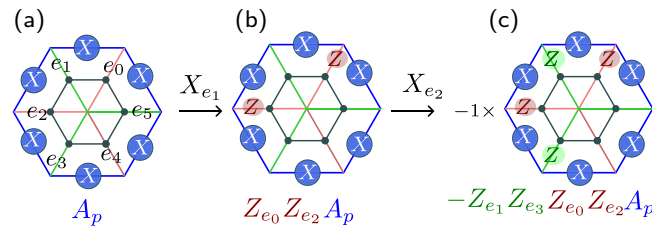


FIG. 10. The modification of plaquette stabilizer A_p after repeated application of X on its interior edges. (a $\rightarrow$ b) Applying X to a green edge e_1 appends the product $Z_{e_0} Z_{e_2}$ of Pauli- Z operators on the two adjacent red edges to the original plaquette operator. (b $\rightarrow$ c) Applying X to a red edge e_2 appends $Z_{e_1} Z_{e_3}$ to the plaquette operator along with a minus sign, due to anticommutation with the existing Z_{e_2} from the previous step. Updating vertex stabilizers, or updating plaquette stabilizers after applying Z to an external edge, changes the resulting stabilizers only up to a sign.

operators either commute or anticommute with the operators B_v and A_p (when acting on the state at that instant of time) that we measure, so standard stabilizer tableau techniques can be used to simulate these measurement processes. Here B_v commutes with all instantaneous stabilizers, and always has a definite measurement outcome given by the sign of the associated instantaneous vertex stabilizer. On the other hand, A_p may either commute or anticommute with a given instantaneous stabilizer. When it fails to commute, the instantaneous stabilizers must be modified in a manner consistent with the probabilistic measurement outcome (see Appendix C). If A_p commutes with all instantaneous stabilizers, the measurement outcome is deterministic, and the set of stabilizers need not be modified. However, it can be the case that A_p commutes with all instantaneous stabilizers, but is *not among* the instantaneous list of stabilizer generators. In this case, this deterministic measurement outcome of A_p must be determined from the existing stabilizer generators; we describe how to do this in Appendix C and also provide additional details about encoding the stabilizers in tableaux that can be simulated efficiently.

V. D_4 STEADY-STATE PHASE DIAGRAM

Having established the local correction protocols, we now present the steady-state phase diagram under the continuous action of correction and heralded noise [see Eq. (34)]. After discussing the flag dynamics and reviewing the diagnostics used to characterize the steady-state order, we present the complete dynamical phase diagram as well as the nature of the transition. Toward the end of this section, we also briefly discuss the dynamics in the presence of imperfectly heralded noise.

A. Dynamical phase diagram of flags

We begin by discussing the steady-state phases of the flags, which undergo a classical dynamics that does not depend on the dynamics of the stabilizer defects. Recall that the active phase of the flags refers to the scenario where the absorbing state is not reached for times that are exponentially long in the system size; in contrast, regimes where the absorbing state $n_X^f = n_Z^f = 1$ is the only possible choice of steady state (i.e., where the absorbing state is reached on a timescale that is independent of system size) lie within the *absorbing phase* of the dynamics.

Figure 1(b) shows the numerically obtained steady-state flag density as a function of the correction rates γ_x and γ_z (for noise rate $\eta = 1$). When both correction rates are sufficiently high, we observe a fully active phase, in which the densities of both X and Z flags remain low out to times that are exponentially long. On the other hand, when both correction rates are low, the steady state belongs to the absorbing phase where all flags are raised, such that

no corrections can take place. These two phases are separated by a partially active phase, in which the density of X flags remains low, but Z flags proliferate and rapidly reach a density $n_Z^f = 1$.

To understand this phase diagram, we begin with the X flags. It is straightforward to see that these undergo dynamics identical to those of the X flags in the toric code, since neither the presence of Z flags nor the Z correction steps affect the procedure used to apply X corrections. In particular, the X flags enter into an absorbing state precisely when the X correction rate is below the threshold value for the toric code, i.e., $\gamma_x \leq \gamma_{\text{TC-X}}^c$. For $\gamma_x > \gamma_{\text{TC-X}}^c$, the X flags remain in their active state.

The dynamics of the Z flags, however, is rather different compared with their behavior in the toric code. There are two reasons behind this. First, the correction for the Z flags [see Eqs. (32) and (33)] can only be carried out when all X flags inside a given plaquette are lowered (i.e., in the $n_X^f = 0$ state). Once the X flags enter their absorbing state, the Z correction protocol becomes inert and Z flags also necessarily enter their absorbing state, irrespective of the correction value γ_z . Thus, when $\gamma_x \leq \gamma_{\text{TC-X}}^c$ the system is in the fully absorbing phase irrespective of γ_z . This contrasts

sharply with the case of the toric code, where we observed a partially active phase where X flags have proliferated but Z flags have not.

If the rate γ_x is increased beyond $\gamma_{\text{TC-X}}^c$, the X flags enter an active state. For small values of γ_z , the Z flags remain in the absorbing state in both the toric code and D_4 models, resulting in a partially active phase. However, the threshold value of γ_z required to pass from this partially active phase to the fully active phase is much higher here than in the toric code, and also depends on the value of γ_x . This is due to two key differences between the dynamics of Z flags in the D_4 model relative to the toric code: first, even a low density of X flags can significantly lower the actual rate at which Z corrections are applied; and, second, removing X flags introduces new Z flags in the D_4 setup. In Appendix D, we also provide a mean-field phase diagram, which is obtained by ignoring all intersite flag correlations: this approximation illustrates the effect of the correlations between the locations of the X and Z erasures, which is absent for the toric code dynamics, on the phase diagram.

Figure 11 shows the combined evolution of flag densities, stabilizer defect densities, and logical fidelity at a representative parameter value within each distinct

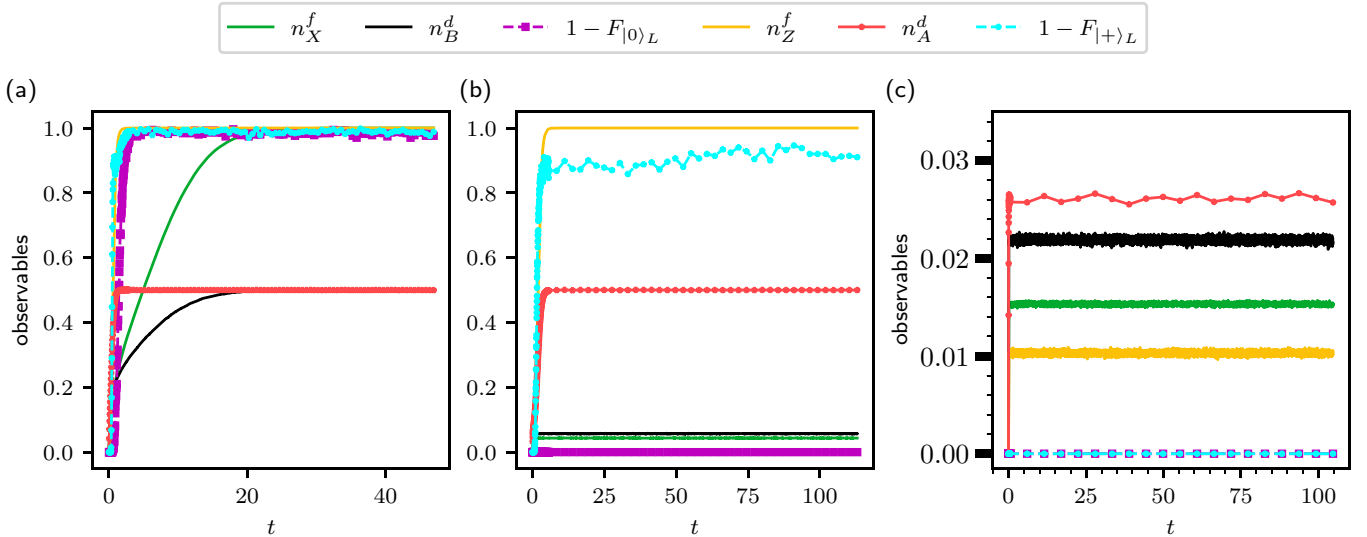


FIG. 11. Time evolution for the D_4 model on three 30×30 sized interpenetrating honeycomb lattices at three representative points in the γ_x - γ_z phase diagram [see Fig. 1(b)]. Plots show the densities of X and Z flags, stabilizer defect densities n_B^d and n_A^d , and the logical error rate $1 - F_{|\tilde{0}\rangle_L(|\tilde{+}\rangle_L)}$ [see Eq. (42)] as a function of time. Time is measured in units of inverse erasure noise rate, set to $\eta = 1$. The flags are initialized to $n_X^f = n_Z^f = 0$ and the initial state of qubits is set to $|\tilde{0}\rangle_L$, except when calculating $1 - F_{|\tilde{+}\rangle_L}$, where it is initialized in the $|\tilde{+}\rangle_L$ state. Each data point is obtained by averaging over 500 independent Monte Carlo trajectories. (a) Absorbing phase ($\gamma_x = 4, \gamma_z = 500$): when the X correction rate is small, X flags proliferate, which also renders Z correction ineffective [see Eqs. (32) and (33)]. Since all flags are trapped in the absorbing state, the defect densities saturate to $1/2$, and all logical information is lost, as witnessed by $F \rightarrow 0$. (b) Partially X -active phase ($\gamma_x = 14, \gamma_z = 500$): while the densities of X flags and vertex defects saturate to small values, Z flags still proliferate. This phase can only protect classical information, as seen by perfect recovery of $|\tilde{0}\rangle_L$ state (magenta curve) but failure to preserve $|\tilde{+}\rangle_L$ state (cyan curve). (c) Active phase ($\gamma_x = 35, \gamma_z = 500$): when the X correction rate is sufficiently large, the long-time value of n_X^f (and of n_B^d) becomes small, allowing Z correction to become sufficiently effective to stabilize n_Z^f and n_A^d densities at small values. In this phase, the logical error probability is close to 0, witnessed by $(1 - F) \approx 0$ for both $|\tilde{0}\rangle_L$ and $|\tilde{+}\rangle_L$ initial states leading to protected quantum information up to long times.

steady-state phase shown in Fig. 1(b). We see that in each case, the dynamics of stabilizer defects closely tracks that of the flags, with both steady-state values and the timescales at which these reach their long-time values closely correlated. Here, the densities of the vertex and plaquette stabilizer defects are defined, respectively, as

$$n_B^d = \frac{1}{2L^2} \sum_v \frac{(1 - \langle B_v \rangle)}{2}, \quad n_A^d = \frac{1}{L^2} \sum_p \frac{(1 - \langle A_p \rangle)}{2}, \quad (40)$$

where summation runs over all vertices v and plaquettes p . Figure 11(a) shows a representative point in the absorbing state, where both flag densities increase rapidly to their steady-state value of 1 (even though this increase occurs more slowly for X flags, as is apparent in the figure). The densities of plaquette defects (created by Z errors) and vertex defects (created by X errors) closely track the respective flag densities and rapidly equilibrate to their steady-state values of $1/2$, indicating that the qubits quickly reach the maximally mixed state.

In the partially active phase, shown in Fig. 11(b), Z flags proliferate rapidly and attain a density of 1, but X flags remain at low density in the steady state. Consequently, the density of plaquette stabilizer defects (caused by Z -type error strings) rapidly saturates to the maximally mixed value $n_A^d = 1/2$, but the density of vertex stabilizer defects remains low in the steady state (indicating a low density of X -type error strings).

Finally, the time evolution for the system in the fully active phase is shown in Fig. 11(c). Here, we see that both flag densities rapidly approach their steady-state values, but, in contrast to the absorbing state, these are now small. As a result, the density of stabilizer defects is also small in the steady state. This low density of stabilizer defects strongly suggests that the steady state has D_4 TO. In the following section, we show explicitly that this is indeed the case.

B. Diagnostics for topologically ordered steady states

Before we present our numerical results for the dynamics of stabilizer defects, we review the diagnostics that we use to characterize the steady-state order. We use two metrics here. The first metric can be viewed as a pragmatic one: we ask whether, after applying a conventional MWPM error-correction protocol, our steady state returns to the original D_4 topological ground state in which it was prepared. The second metric asks whether our steady state can be connected to the defect-free D_4 ground state via a finite-depth quasilocal quantum channel: an affirmative answer would place the steady state in the same mixed-state phase as the pure D_4 ground state [51].

Let us begin with the error-correction metric. The simultaneous $+1$ eigenspace of the A_p and B_v operators of the

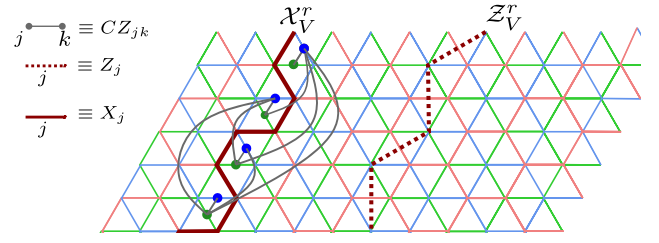


FIG. 12. Logical operators for the D_4 ground space: a 12×6 sized patch of D_4 model with PBCs is shown. The Z logical operator Z_V^r acts along a noncontractible loop on the red dual lattice in the vertical direction. The logical- X operator X_V^r is constructed by applying X along a noncontractible loop C_r on the red honeycomb lattice, followed by a set of CZ gates that couple all blue and green edges in a specific pattern [see Eq. (27)]. Here, starting from the top, every blue edge that is incident to a vertex along C_r is connected to all green edges below it along the path C_r . The remaining logical operators can be defined analogously.

D_4 model is 22-fold degenerate on the torus, reflecting the number of distinct superselection sectors (anyons) of D_4 TO. The states in this subspace can be distinguished based on the values of certain noncontractible loop operators. One choice is to label the 22 ground states according to their eigenvalues under the noncontractible Abelian loop operators, given by $Z_{H/V}^c = \prod_{k \in l_{H/V}} Z_k = \pm 1$, where l_H (l_V) are horizontal (vertical) loops on the dual of the c -colored honeycomb lattice, as shown in Fig. 12. We term these the “logical- Z ” operators. While there are 2^6 different choices for the eigenvalues of these logical operators, only 22 of them are consistent with $A_p = B_v = 1$ for all p, v [63]. The operators that switch between the allowed logical- Z eigenstates are the noncontractible string operators associated with transporting non-Abelian anyons around the torus. We refer to these as the “logical- X ” operators, X_H^c, X_V^c . As shown in Fig. 12, these non-Abelian string operators consist of a string of X operators decorated with nonlocal CZ gates. A second basis choice for the 22 locally indistinguishable ground states is by their values under the vertical Z_V and X_V logical operators.

In order to analyze the fate of the encoded logical information in the ground space of the D_4 model under the noise and correction dynamics, it will be useful to consider two different initial logical states, one in each of these basis choices:

$$\begin{aligned} |\tilde{0}\rangle_L &:= \bigotimes_{p_c, v_c, c} |A_{p_c} = B_{v_c} = +1; Z_H^c = Z_V^c = +1\rangle, \\ |\tilde{+}\rangle_L &:= \bigotimes_{p_c, v_c, c} |A_{p_c} = B_{v_c} = +1; X_V^c = Z_V^c = +1\rangle. \end{aligned} \quad (41)$$

The ability to recover *both* of these states from the time-evolved noisy-mixed state is a measure indicating that both the amplitude and the relative phase amongst the logical-computational states can be protected.

To determine whether this logical information is preserved in the steady state, we use a MWPM decoder appropriate to the D_4 TO. At any given instant, the state of the system is represented by a product of vertex eigenstates and, potentially, a superpositions of plaquette eigenstates. The decoder acts by first measuring all vertex stabilizers, and pairing up the vertex defects using MWPM. Pairs of vertex defects are eliminated by applying strings of Pauli- X operators to obtain a state with $B_v = +1$ at every vertex. As described previously, this can create further plaquette (A_p) defects. Since all A_p operators commute in the $B_v \equiv +1$ state, we can next measure all plaquette operators, collapsing the state onto a fixed configuration of plaquette defects. The plaquette defects are then paired up using a MWPM decoder on the dual lattice, and eliminated by applying Pauli- Z operators along the appropriate dual strings. We compute the failure of this decoding process by evaluating the fidelity of the resulting decoded state $|\psi_{\text{decoded}}^{(r)}\rangle$ with the initial state:

$$1 - F_\Phi = 1 - \frac{1}{N_r} \sum_{r=1}^{N_r} |\langle \psi_{\text{decoded}}^{(r)} | \Phi \rangle|^2, \quad (42)$$

where Φ is one of the initial code states defined in Eq. (41) and the overlap is averaged over N_r independent Monte Carlo realizations.

Our second diagnostic utilizes the two-way channel connectivity equivalence relation placed on the space of mixed states [51]: two states ρ_1 and ρ_2 belong to the same mixed-state phase if there exist a pair of quasilocal, finite-depth quantum channels Σ_{12}, Σ_{21} such that $\Sigma_{21}(\rho_1) = \rho_2$ and $\Sigma_{12}(\rho_2) = \rho_1$. In our case, the challenge is to determine whether there exists a finite-depth quantum channel that takes the steady-state density matrix ρ_{ss} back to the initial error-free state ρ_1 . We assess this by measuring the time required for the steady state to evolve back to the error-free state under the Lindbladian equation (34) in the absence of noise, i.e., under the action of only the error-correction terms. Specifically, we examine how the time t at which $\rho_1 \sim e^{\mathcal{L}_{\eta=0}t} \rho_{ss}$ scales with system size, with $\mathcal{L}_{\eta=0}$ given by Eq. (34) with $\eta = 0$. The density matrices ρ_1 and ρ_{ss} are considered as being connected via a finite-depth quantum channel as long as this time grows at most (poly)logarithmically with L [51]. Here, two mixed states are considered to be close in terms of the fidelity, i.e., $\rho \sim \rho'$, if $F(\rho, \rho') > 1 - \varepsilon$ for some $\varepsilon > 0$.

We emphasize that this is, in principle, a different (and stronger) criterion than requiring that the logical information encoded in the initial state can be recovered, since our MWPM decoder is highly nonlocal and need not guarantee the existence of a pair of quasilocal finite-depth two-way channels. Interestingly, however, at least in the context of the toric code, Ref. [51] showed that the two criteria are equivalent in practice.

C. Phase diagram of the D_4 model

We now apply the criteria outlined in the previous subsection to show that, in the D_4 model with perfectly heralded errors, (a) the flag dynamics fully dictates the steady-state order as measured by MWPM decodability, and that (b) the active phase has non-Abelian steady-state TO as measured by the criterion of two-way connectivity.

We begin by discussing the correspondence between flag dynamics and the logical fidelity (42) shown in Fig. 11 [also see Fig. 1(b)]. When both X and Z flags are in the absorbing state, as noted previously, the densities of both plaquette and vertex defects closely track the respective flag densities and rapidly equilibrate to their steady-state values of $1/2$, indicating a fully decohered mixed state. This proliferation of stabilizer defects leads to a rapid decay of the logical fidelities for both the $|\tilde{0}\rangle_L$ and $|\tilde{+}\rangle_L$ states.

In the fully active phase, Fig. 11(c), where the density of both flags and stabilizer defects is small in the steady state, we expect that the strings of errors that create stabilizer defects (which must lie along strings of flags) have a very high probability of being short. This is clearly borne out in our numerics, as seen in Fig. 11(c): the probability of decoding to the incorrect logical state is vanishingly small (i.e., the logical fidelity is essentially 1 after decoding the steady state, for all times shown). In other words, from the perspective of decodability, our fully active phase has non-Abelian D_4 TO.

Figure 11(b) shows results for the partially active regime, where the steady-state density of plaquette stabilizer defects is $1/2$, but the density of vertex stabilizer defects remains low in the steady state (indicating a low density of X -type error strings). In this phase, we see a distinction between the fidelities of the MWPM-decoded state and the initial state, depending on which of the two logical states we use. Specifically, because vertex defects are rare, the probability of acting with a noncontractible loop of X strings (and, hence, a noncontractible non-Abelian loop operator, after the full error-correction protocol is executed) during the MWPM decoding step is small. This leads to a very small probability of the logical errors in $|\tilde{0}\rangle_L$ state that is sensitive only to long strings involving X operators. Logical- X information, on the other hand, is sensitive both to non-Abelian strings and the Abelian Z strings; in this partially active phase, where strings of Z errors proliferate, this information is rapidly lost. In other words, the partially active phase is a good classical memory, similar to the partially active phase in the toric code, but fails to preserve the full quantum information of the D_4 state.

Next, we turn to our second criterion for examining the phase diagram, based on the existence of two-way quantum channel connectivity between the initial state and the steady state. Figure 13 shows the fidelity $F(t) = \langle \psi_{D_4} | e^{\mathcal{L}_{\eta=0}t} [\rho_{ss}] | \psi_{D_4} \rangle$ of the initial state with the

time-evolved steady state ρ_{ss} for system sizes ranging from 24×24 sites to 144×144 sites. (Here $\mathcal{L}_{\eta=0}$ indicates time evolution under the correction-only Lindbladian.) To allow simulations out to larger system sizes, we simulate only the flag dynamics, and plot the quantity

$$F_{\text{flag}}(t) := \langle n_X^f = n_Z^f = 0 | \rho_{\text{flag}}(t) | n_X^f = 0, n_Z^f = 0 \rangle, \quad (43)$$

which estimates the logical fidelity at time t from the flag density. Here, $|n_{X,Z}^f = 0\rangle$ is state where both X and Z flags on all sites are set to 0 and ρ_{flag} is the reduced density matrix of the flags. This upper bounds the full fidelity, since a state with no flags automatically has no defects or error strings, and thus will error-correct into the correct logical state. We verify in Fig. 13(b) that for $L \leq 30$, where we can perform the full stabilizer simulation, this is a good approximation for the full fidelity.

As seen in Fig. 13(c), the resulting fidelities are consistent with a local correction time that scales logarithmically with the system size, providing compelling evidence that our steady state is in the same mixed-state phase as the pure D_4 topologically ordered state. While the timescale for the state to get close to the initial flag-free state in terms of fidelity increases logarithmically with increasing L , the local densities of flags approach zero in system-size-independent time. Since correction can only remove flags from the boundaries of flag clusters, the timescale to recover the initial state should be determined by the size of the largest connected clusters of flags. In Appendix E, we show that this is indeed the case: we plot the distribution of flag clusters as a function of their size, and find that the average size of the largest cluster increases logarithmically with increasing system size. We also expect that the partially active phase we have found here represents an intrinsically mixed-state TO (distinct from the intrinsically mixed TO characterizing the active phases of the toric code) since it encodes a nontrivial classical memory (see Refs. [37,53] for a discussion of such intrinsically mixed TO phases).

D. First-order transitions

Finally, we study the nature of the transitions separating these different steady-state phases, which we find to be first order. In the previous section, we established that the active to absorbing state transition of the flags coincides precisely with the transition of the stabilizer system. As it is numerically less expensive to simulate the reduced dynamics of flags, here we exploit the aforementioned fact and present numerical results based on simulations of only the flag densities, allowing access to higher system sizes.

We emphasize that the flags are classical degrees of freedom and undergo a *classical* nonequilibrium absorbing state transition [77,78,110]. The density of active sites and the survival probability of the initial activity serve

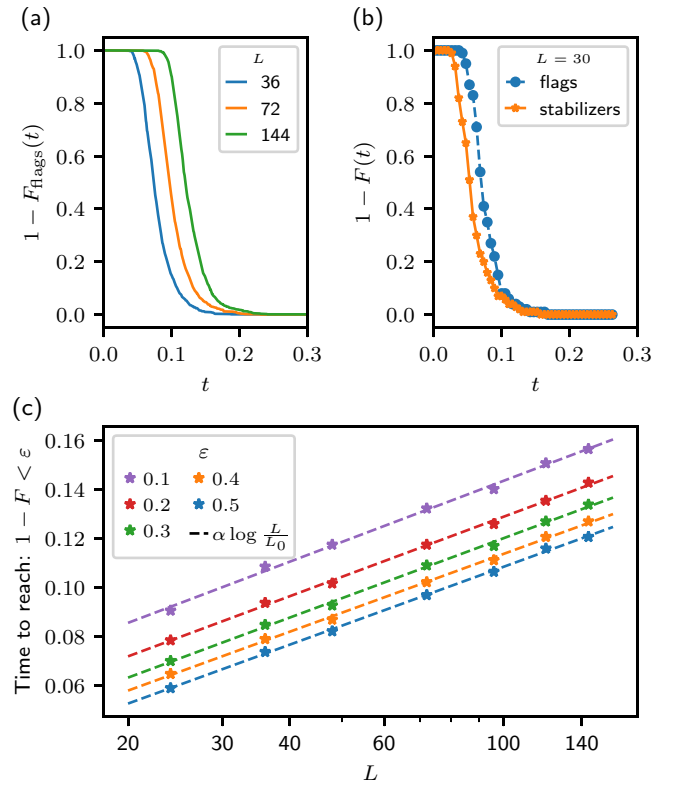


FIG. 13. Two-way connectivity in the D_4 active phase: we estimate the time required for noiseless dynamics $\eta = 0$ to take the (quasi)steady state ρ in the active phase ($\gamma_x = 35$, $\gamma_z = 500$ simulated up to 500 Monte Carlo sweeps) back to the pure D_4 ground state $|\psi_0\rangle$. To numerically access the larger system sizes, we estimate the fidelity of the time-evolved state based only on the flag densities using F_{flags} [see Eq. (43)]: (a) $1 - F_{\text{flags}}(t)$ evaluated for various system sizes with linear dimension L ; data obtained by averaging over 1000 independent Monte Carlo realizations; (b) $1 - F_{\text{flags}}(t)$ (blue) and $1 - F_{\text{full}}(t)$ of the qubit system (orange) are plotted as a function of time for $L = 30$. This demonstrates that $F_{\text{flags}}(t) < F_{\text{full}}(t)$ and, hence, recovery timescales obtained using F_{flags} upper bound the recovery time computed in terms of the full fidelity of the stabilizer system. (c) Time τ to achieve $1 - F_{\text{flag}}(t) < \epsilon$ is plotted as a function of system size L for various cutoff values ϵ on a linear-log-scale plot. The dotted lines show best fits to the functional form $\alpha \log(L/L_0)$. We note that the presence of even a single flag in a given trajectory leads to zero fidelity with $|n_{X,Z}^f = 0\rangle$ state.

as order parameters for such transitions. The active phase corresponds to a nonvanishing probability of persistent activity out to times exponentially long in the system size, whereas the system becomes inactive after a finite time in the absorbing phase. The location of this transition can be determined from the timescale after which activity disappears. In the present case, this can be determined from the time it takes for the density of active (i.e., unflagged) sites, given by $1 - n_X^f$ or $1 - n_Z^f$, to fall to small values: this changes from exponential in system size in the active

phase, to a constant value independent of system size in the absorbing phase.

As in the equilibrium case, these transitions may exhibit a first-order or second-order character. However, unlike in the equilibrium case, there is no notion of free energy for such absorbing state transitions [110]. Nonetheless, we can still make an analogy to the equilibrium scenario by monitoring the behavior of order parameters. For instance, when the system undergoes a continuous transition (e.g., in the case of directed percolation), the spatial and temporal correlation lengths diverge at the critical point [78]. This results in the power-law decay of density of active sites at the critical point, which can be analyzed using standard scaling techniques analogous to equilibrium critical phenomena [77]. As we show in the following though, near the transition point of the models considered here, a significant fraction of Monte Carlo trajectories enter the absorbing state while the remaining trajectories maintain an active configuration. Such bistable behavior is instead more analogous to first-order equilibrium transitions [111]. Analysis of the discontinuous transitions to the absorbing state using the finite-size scaling methods generally involves modifications to the dynamics itself [112], which we do not pursue here. Instead, we observe that the system exhibits a bistability of both the absorbing and the active (with much lower flag density) configurations, as also observed in various discontinuous absorbing phase transitions [77, 113, 114], and qualitatively resembling equilibrium first-order transitions.

To study the nature of the transition, we begin by locating it precisely. Here, we determine the location of the transition along the $\gamma_z = 500$ cut in the phase diagram shown in Fig. 1(b). For this, we exploit the fact that the time τ to reach a high flag density (in this case, 0.65) grows exponentially in system size in the active phase, while in the absorbing phase it reaches an L -independent value for modest system sizes. We can thus use the change in the concavity of the τ vs L curve on a log-log scale to pinpoint the transition from the absorbing to the partially X -active phase. We note that the dynamics of X flags on the honeycomb lattice of a fixed color is independent of X flags on the remaining lattices and all of the Z flags. Hence, we locate this transition by only simulating the X flags on a single honeycomb lattice, which allows access to higher system sizes. In Fig. 14(a), we see that for $\eta = 1$, $\gamma_z = 500$, this happens at $\gamma_x = 6.45 \pm 0.03$ independent of the value of γ_z . Similarly, the transition from the partially active into the fully active state is observed at $\eta = 1$, $\gamma_z = 500$, $\gamma_x = 18.25 \pm 0.25$, as shown in Fig. 14(b).

Having located the two transitions precisely, we now study the nature of the transition by monitoring the density of the X and Z flags as we cross each transition point. The probability distribution of the density of X flags at the first transition point ($\gamma_x = 6.45$) is shown in Fig. 15(b). The snapshot is taken at a time t such that $\langle n_X^f(t) \rangle_{\text{avg}} = 0.6$.

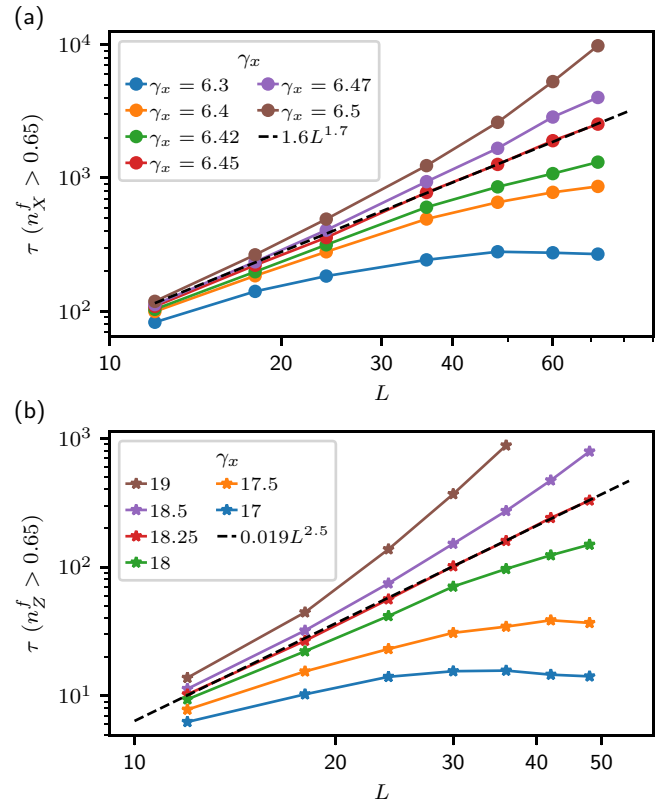


FIG. 14. Determination of the first-order transition point based on the decay of flag densities toward the absorbing state in a finite-sized system: the time τ required for the erasure flag density to cross a finite constant value, chosen to be $n_0^f = 0.65 < 1$, when initialized in a state with no flags, is plotted as a function of linear dimension L for varying values of X correction rates γ_x close to the transition. The critical transition rate is characterized by the value of γ_x where the curve changes from concave up (active phase for infinite system) to concave down (absorbing phase for infinite system). (a) Time τ for n_X^f to reach 0.65 as a function of system size for parameters near the transition from absorbing to partially X -active (where Z flags are in their absorbing state) phase. Because the X flags on a given honeycomb lattice evolve independently of the Z flags (and the X flags of the remaining two lattices), we only simulate X flags of a single honeycomb lattice, allowing access to larger system sizes. Each data point is obtained by averaging over 1000 Monte Carlo realizations. (b) Time τ associated with the Z flag density is shown as a function of system size for $\gamma_z = 500$ and varying values of γ_x close to the transition from partially X -active to fully active phase. The data points are obtained by averaging over 500 Monte Carlo realizations.

We observe a bimodal structure: in a significant fraction of samples, the system has maintained a density of flags that is much smaller than 1, leading to a peak at a low flag density $n_X^f \approx 0.2$; in the remaining samples, the system is already in the absorbing state, leading to a δ -function peak at $n_X^f = 1$. These two peaks are well separated in the space of flag densities, indicating that the probability of being

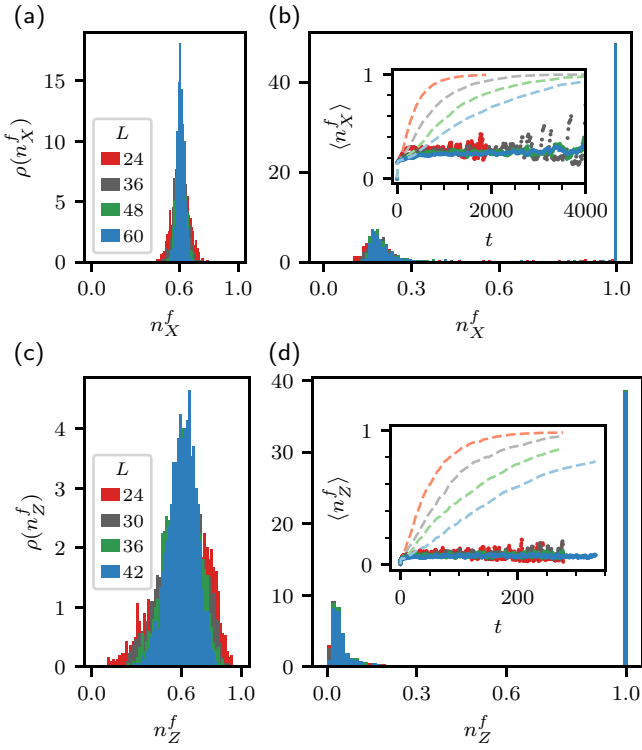


FIG. 15. Bistability at the transitions: the probability distribution function $\rho(n_{X(Z)}^f)$ of the density of the number of $X(Z)$ flags ($n_{X(Z)}^f$) among Monte Carlo trajectories at time t_* . The time t_* is chosen such that the average density reaches an arbitrary finite value $n_{\text{avg}}^f = 0.6$. The data are shown for 500–2000 independent Monte Carlo realizations. (a),(b) The transition from absorbing to partially active state is analyzed by monitoring the density of X flags on one of the honeycomb lattices of the D_4 model. (a) Absorbing phase (shown for $\gamma_x = 4$, independent of γ_z): the distribution has a single peak centered near the average value and it moves toward $n_X^f = 1$. (b) Near the absorbing to partially active transition ($\gamma_x = 6.45$, independent of γ_z): the distribution is bimodal where some of the Monte Carlo trajectories enter the absorbing state ($n_X^f = 1$) but a finite fraction of trajectories remain active with a small density of X flags, witnessed by a distinct secondary peak at $n_X^f < 0.5$. The inset shows the full n_X^f (dashed lines) and the average density of flags conditioned on Monte Carlo realizations still in the active state (solid dots) as a function of time. (c),(d) Similar results are shown for the transition from X -active to fully active at $\gamma_z = 500$, controlled by the density of Z flags. (c) Unimodal distribution of Z flag density changes into (d) bimodal distribution at $\gamma_x = 18.25$.

in a state with an intermediate flag density is vanishingly small. As discussed previously, such bistable behavior is a hallmark of first-order transitions, where the order parameter changes discontinuously by shifting the relative weights of two distinct configurations [111,112].

The bistability is observed for a finite window of time, after which most of the trajectories are trapped in the

absorbing state due to finite size effects. The time dependence of the average flag density shown in Fig. 14 suggests that this timescale grows polynomially with L . To see that this is indeed the case, in the inset of Fig. 15(b) we show the density of X flags computed by averaging only over those realizations that have not entered the absorbing state at a given time. This corresponds to the density of flags contained in the realizations near the secondary peak in Fig. 15(b). We observe that while the average density decays as a function of time, the conditioned density remains stable. The density among the surviving trajectories remains at a fixed value for a long time, indicating the robustness of the peak at low flag density $\langle n_X^f \rangle = n_0$. This demonstrates that (given the relatively narrow peak widths) we can approximate the total density as a weighted average

$$n_X^f(t) = P_{\text{absorb}}(t) + n_0(1 - P_{\text{absorb}}(t)), \quad (44)$$

where $P_{\text{absorb}}(t)$ is the fraction of trajectories that are in the fully occupied absorbing state. Thus, the timescale over which bistable behavior is observed (i.e., $P_{\text{absorb}} < 1$) for a given system size is the same as τ estimated from the full densities in Fig. 14; that is, the time window grows as $\text{poly}(L)$ near the transition. In contrast, for γ_x values deep in the active phase, the time to reach an absorbing configuration grows exponentially with system size.

We note that this bimodality is sharply distinct from the time evolution seen for systems in the absorbing phase. In this case, at any given time, the probability distribution is characterized by a single peak, which moves to the right (i.e., to higher n_X^f) with increasing time, until all the weight is concentrated on the absorbing state. Figure 15(a) shows this for $\gamma_x = 4$ (in the absorbing phase) evaluated at the time t' ($t' \neq t$), for which the flag density $\langle n_X^f(t') \rangle_{\text{avg}} = 0.6$ is equal to that shown in Fig. 15(b).

Figures 15(c) and 15(d) show similar data for the densities of Z flags near the transition from the partially active phase (which is an absorbing state for the Z flags) into the fully active phase. Again, bistable behavior is observed near the transition [Fig. 15(d)], whereas in the Z -absorbing phase, a single peak is observed about the mean flag density, which moves to the right as time increases until all the weight is in the Z absorbing state $\langle n_Z^f \rangle = 1$.

E. Imperfect heralding

Thus far, we have only considered the situation where all error processes are detectable and can be perfectly converted into erasures by raising the flags. Under such idealized conditions, we demonstrated that perfect heralding of errors provides local information that is sufficient for correcting them and stabilizing a topologically ordered steady state. In practice, however, some of the errors are

not converted into erasures due to experimental imperfections [64]. Moreover, qubits are also affected by error mechanisms that do not take them out of the computational subspace; these events cannot be detected by measurement without disturbing the state of qubit [58,59]. Here, we investigate the effect of introducing a low rate of unheralded errors, as will realistically be the case for even the most promising erasure-qubit platforms.

The specific nature of these unheralded noise processes depends on the details of the qubit platform. Here, we phenomenologically model them as an uncorrelated depolarizing channel acting on qubits without raising any of the erasure flags. Let ϕ_e be the fraction of errors that are heralded by flags and converted into erasures (i.e., $\phi_e = 1$ corresponds to the perfect heralding situation being considered so far). Then, the Lindblad superoperator for unheralded error is given by

$$\mathcal{L}_{\eta_p}(\rho) = \frac{\eta(1 - \phi_e)}{3} \sum_j (X_j \rho X_j + Y_j \rho Y_j + Z_j \rho Z_j - 3\rho), \quad (45)$$

where j runs over all qubits in the system. The heralded fraction of the noise is represented by the erasure noise model in Eq. (6), where we replace $\eta \rightarrow \eta\phi_e$. The time evolution of the state under the combined heralded and unheralded noise, along with correction operations, is generated by $\mathcal{L}_{\eta_e} + \mathcal{L}_{\eta_p} + \mathcal{L}_C$. The unheralded errors pair-create stabilizer defects that are not connected by strings of flags. Our correction protocol does not act on such defects if there are no flags in their vicinity. Hence, once such defects are created their evolution is effectively uncorrelated from that of their partner. In a long-time limit, for any finite value of unheralded error rate, this results in a finite density of defects that are well-separated from their partners. We expect that upon pairing these defects using MWPM decoder, we will not get back to the initial state with certainty, meaning that our steady state has lost its topological properties. In Figs. 16(a) and 16(b), we demonstrate this by evaluating the fidelity of the state obtained by decoding the time-evolved state using MWPM decoder with respect to two different initial states in the D_4 ground-state space. The fidelity in the steady state vanishes for all values of erasure fraction $\phi_e < 1$.

While the steady-state quantum memory is unstable against unheralded errors, however, when most of the errors are heralded, the correction protocol nevertheless results in a significant enhancement of the timescale up to which the encoded information can be recovered. Specifically, the time required for the fidelity to decay to a given value increases with increasing value of the heralding fraction ϕ_e as $\tau \sim (1 - \phi_e)^{-\alpha}$, where numerically we observe that $\alpha \approx 0.5$.

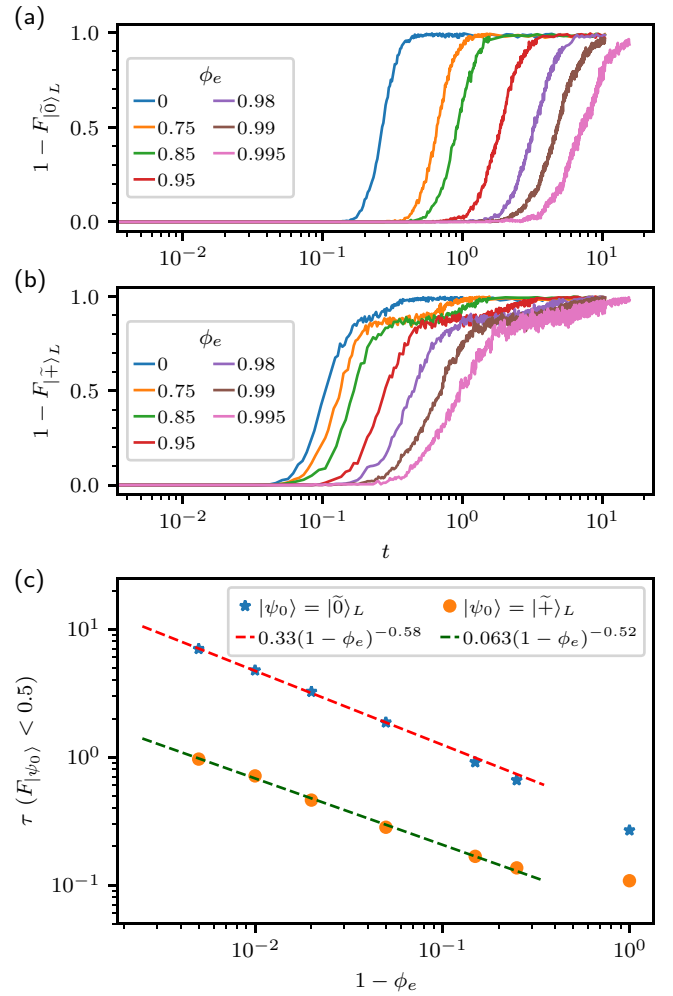


FIG. 16. D_4 model in the presence of unheralded noise: the logical error probability $1 - F$ of a system affected by a combination of heralded (rate ϕ_e) and unheralded (rate $(1 - \phi_e)$) noise, along with the correction channel, is plotted as a function of time for varying values of ϕ_e . Here, F_{ψ_0} is the fidelity of the initial logical state $|\psi_0\rangle$ with the decoded state. Data are shown for $L = 18$ sized system, averaged over 250–500 Monte Carlo realizations where all flags are initialized in the $n_{X_j}^f = n_{Z_j}^f = 0$ state. The logical error probabilities (computed using the MWPM decoder) are shown for qubits initialized in (a) $|\tilde{0}\rangle_L$ and (b) $|\tilde{+}\rangle_L$ logical states. (c) The time τ required for the fidelity to decrease below 1/2 is plotted as a function of the unheralded noise fraction. The dashed lines are obtained by fitting the data points (except $\phi_e = 0$) to a power law $\tau = a(1 - \phi_e)^b$.

VI. CONCLUSIONS

In this paper, we have presented a fully local dynamical protocol that, under the assumption of perfectly heralded noise, can stabilize steady-state phases with Abelian ($\mathbb{Z}_2$) or non-Abelian (D_4) TO in two dimensions. Building on our previous work in one dimension [40], here we have demonstrated that the information about error locations

provided by erasure flags can be exploited to effectively confine the errors, leading to a nontrivially ordered active dynamical phase for noise rates below a certain threshold. After illustrating our correction protocol for the toric code, we have used three different metrics to show that the active phase in the D_4 model constitutes a legitimate dynamical phase of matter. First, we have verified that this phase effectively preserves quantum information, in the sense that the logical information can be faithfully recovered by applying the MWPM decoder. Second, we have provided compelling numerical evidence that this active phase lies in the same mixed-state phase as the pure D_4 ground state: specifically, we show that a local quantum channel can take a state in this active phase and return it to a pure D_4 ground state in a time that scales at most logarithmically in the system size. Finally, we have shown that a first-order phase transition separates the dynamics of flag and defect distributions in this active phase from neighboring dynamical regimes. In addition to the fully decohered absorbing phase, we have also established the existence of an intermediate steady-state phase, wherein classical (but not quantum) information can be preserved with high fidelity.

Several aspects of our results are worth emphasizing. First, we have tested our model in the presence of *unbiased* heralded noise, even though heralded errors most often occur in systems where the noise is highly biased. As such, unbiased noise presents the most challenging obstacle for our non-Abelian decoder, since there is a high probability that any given application of the noise channel creates new non-Abelian defects. In a scenario where the noise is biased, such that Z -type errors occur more frequently than X -type errors, the majority of the errors created will be Abelian plaquette defects, which require considerably less overhead to correct. Under these conditions, the Z correction rate required to maintain a fully active D_4 phase would be much closer to the values seen for the toric code than those observed for the D_4 model with unbiased noise.

Second, our local decoder exploits the fact that the fusion rules of the D_4 TO are acyclic, meaning that two non-Abelian anyons of the same type fuse only to Abelian anyons. This property allows us to remove non-Abelian defects by applying a sequence of single-site X operators. While advantageous for a general non-Abelian decoder (such as the MWPM protocol used in Sec. IV), this feature is essential for our local decoder, since the string operators that pair-annihilate non-Abelian anyons *without* generating further excitations along their length are high-depth unitary circuits that cannot be decomposed into a sequence of purely local operations applied independent of the correction history. In replacing these high-depth unitary circuits with strings of onsite X operators, the main price that we pay is the creation of large numbers of Abelian anyons along the strings' length; however, these can be corrected

locally at sufficiently high Z -correction rates. In principle, we expect that our correction protocol can be generalized to other acyclic anyon theories, e.g., those described by $\mathcal{D}(G)$, where G is some class- n nilpotency group (where n is finite). It would be interesting to investigate the conditions under which these generalizations also admit fully active phases that remain topologically ordered, leading to self-correcting quantum memories when the noise is perfectly heralded. It remains an open question whether our protocol generalizes to the case when G is solvable but not nilpotent (e.g., S_3), where state preparation remains efficient under measurement and feedforward [85]; this question is particularly interesting since anyon theories based on solvable but not nilpotent groups G can be used for universal topological quantum computation [94,115].

Third, in realistic experiments (amongst other imperfections), the measurement-plus-feedback process will itself be error-prone, and it is natural to wonder how our protocol fares in this context. As exhibited by the robustness of the fully active phase to X correction processes, which introduce additional heralded errors, these new types of errors are not expected to qualitatively alter the phase diagram, provided that the errors remain perfectly heralded. However, this is not a realistic expectation for current experiments; instead, we conclude that measurement failures are likely to introduce unheralded errors, which will destroy the steady-state order and lead instead to a finite lifetime for the topological memory. Nonetheless, as we show in Fig. 16, our correction protocol enhances the lifetime of the encoded quantum information when the rate of unheralded errors remains low.

Our results also provide insight into steady-state phases for models in a TO phase away from their fixed points (such as the toric code or D_4 model discussed here). The ground state of any point in the TO phase is connected to the respective fixed-point ground state by finite-depth quasilocal unitary operators [1,116,117]. This, in turn, means that the code space of such a general point in the TO phase will be defined by quasilocal quasistabilizers. As a result, single-qubit errors will create more quasistabilizer violations compared with the fixed-point state. However, if these errors are perfectly heralded, then the resulting quasistabilizer violations will still be connected by strings of flags on some coarse-grained level. We expect that an appropriately modified version of our protocol can successfully target a stable active phase even in this case.

From a practical point of view, we also briefly commented on the feasibility of achieving the correction rates necessary to stabilize the TO phase on current hardware. It turns out that the quantum memory phase requires a relatively large rate $\gamma/\eta \approx 500$ of correction processes involving entangling gates. In practice, however, the stabilizer measurements in our protocol have to be performed

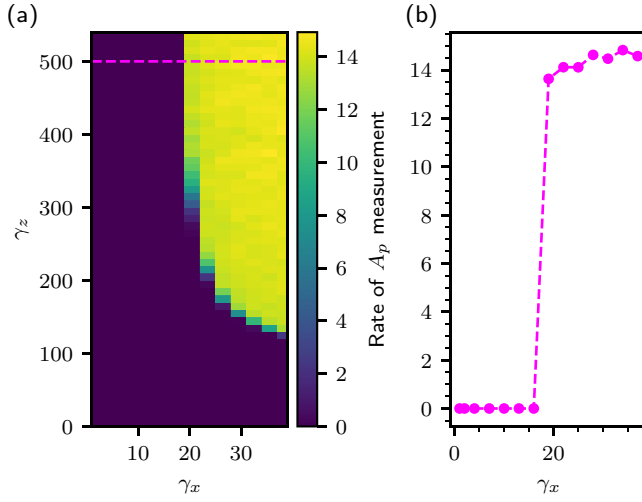


FIG. 17. (a) The rate at which plaquette stabilizers A_p are measured in the steady state is shown for perfectly heralded dynamics in the γ_x – γ_z parameter space of the D_4 model. We compute this rate in the Monte Carlo simulation of flags by counting the average number of A_p measurements per plaquette proposed by the correction jump operators in a single Monte Carlo sweep. Data are shown for an $L = 48$ sized system at a time $t_f = 20$ and averaged over 100 Monte Carlo realizations. The measurement rate goes to zero in the absorbing and partially active phase (purple region) since all Z flags are raised in this regime. In the active phase (yellow region), the plaquette measurements are applied at a finite rate by the correction channel. (b) The measurement rate is shown for a fixed value of Z correction rate $\gamma_z = 500$ as a function of X correction rate [horizontal cut along dashed line in (a)]. The measurement rate jumps from zero to a finite value near the transition where the steady-state density of Z flags changes discontinuously.

only when the classical erasure flags are in an appropriate correctable configuration. Thus, the actual rate at which the entangling gates are implemented, particularly during Z -correction protocol, is expected to be lower than the bare correction rate γ_z . In Fig. 17 we show the rate at which the measurement and feedback operations relevant for Z correction are performed (i.e., measurement of A_p) at various points in the phase diagram, revealing that the actual gate implementation rates are much lower than the bare correction rates throughout the active phase.

Finally, a technical contribution of this work is an extension of the conventional stabilizer tableau to non-Pauli quasistabilizers, which permits efficient simulation of open quantum dynamics for the D_4 model. A necessary ingredient is that the quasistabilizers commute when restricted to the ground-state subspace (i.e., the ground space is frustration-free). Further, the adjoint actions of the noise channels considered here map local operators to local operators. In the future, it would be interesting to understand more generally which properties of the noise and recovery

jump operators are necessary in order for our generalized stabilizer tableau formalism to apply.

ACKNOWLEDGMENTS

A.P. is grateful to Ramanjit Sohal and Nat Tantiwasadakarn for helpful conversations. S.G. thanks Curt von Keyserlingk, Tibor Rakovszky, and Jeff Thompson for helpful discussions. The authors acknowledge helpful conversations with David Huse. F.J.B. and S.C. are grateful for the support of NSF DMR-2313858. S.G. acknowledges funding from an Institute for Robust Quantum Simulation (RQS) seed grant. This material is based upon work supported by the Sivian Fund and the Paul Dirac Fund at the Institute for Advanced Study and the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC0009988 (A.P.). The authors acknowledge the computational resources provided by the Minnesota Supercomputing Institute (MSI) at the University of Minnesota.

APPENDIX A: DETAILS OF NUMERICAL MONTE CARLO SIMULATIONS

In this Appendix, we provide the details of the Monte Carlo algorithm used to simulate the Lindblad dynamics. The system is time evolved using random sequential updates (see, e.g., Ref. [78]) on the state of the flags and the stabilizer tableau. Here, we present a rough outline of the procedure for the local noise-correction dynamics of the D_4 model; the toric code is simulated in an analogous manner.

The flags are initialized in the state $n_X^f = n_Z^f = 0$ for all sites and the stabilizer tableau is initialized in one of the logical states of the D_4 stabilizers (see Sec. IV A). Recall that ϕ_e corresponds to the fraction of heralded errors (i.e., $\phi_e = 1$ refers to perfect heralding) and that γ_x, γ_z parameterize the rate of X and Z correction, respectively. Then, the state is evolved up to time t_f as follows:

```

while  $t \leq t_f$  do
  for  $i = 1, 2, \dots, 3L^2$  do
    sample a random number  $1 \leq j \leq 3L^2$  from
    the uniform distribution to choose the location of the
    qubit on which  $\mathcal{L}_\eta$  acts
    sample a binary random number ( $v_j$  and  $p_j$ ) to
    choose one of the two vertices/plaquettes connected to
    edge  $j$  where  $\mathcal{L}_C$  will act
    sample another random number to choose the
    update rule for the state (with  $c_0 := 1/(\eta + 2\gamma_x/3 +$ 
     $\gamma_z/3)$  here):

```

with probability $c_0\eta\phi_e/4$: Apply X_j and set $n_{X_j}^f = n_{Z_j}^f = 1$
 with probability $c_0\eta\phi_e/4$: apply Z_j and set $n_{X_j}^f = n_{Z_j}^f = 1$
 with probability $c_0\eta\phi_e/4$: apply $Z_j \times X_j$ and set $n_{X_j}^f = n_{Z_j}^f = 1$
 with probability $c_0\eta\phi_e/4$: set $n_{X_j}^f = n_{Z_j}^f = 1$
 with probability $c_0\eta(1 - \phi_e)/3$: apply X_j
 with probability $c_0\eta(1 - \phi_e)/3$: apply Z_j
 with probability $c_0\eta(1 - \phi_e)/3$: apply $Z_j \times X_j$
 with probability $2c_0\gamma_x/3$: apply the appropriate jump operator from the X correction move-set that is consistent with the flag structure at vertex v_j
 with probability $c_0\gamma_z/3$: apply the appropriate jump operator from the Z correction move-set that is consistent with the flag structure at plaquette p_j
end for
 $t \rightarrow t + \delta t := t + \frac{1}{\eta + 2\gamma_x/3 + \gamma_z/3}$
 Record the values of the observables in the state at the end of the Monte Carlo sweep
end while

Throughout this work, we set the rate of noise $\eta = 1$. Note that the number of Monte Carlo sweeps required to simulate the system up to a fixed time t_f increases for increasing values of correction rates. We present additional data in Fig. 18 to show that the final time t_f chosen to compute the phase diagram in Fig. 1 provides a good estimate of the densities in the genuine long-time steady state. In particular, we verify that the variation of the total density over some time interval Δt

$$\delta[n^f] := n^f(t = t_f) - n^f(t = t_f - \Delta t), \quad (\text{A1})$$

where $n^f = \frac{1}{2}(n_X^f + n_Z^f)$, remains small at late times t_f (away from the phase boundaries).

APPENDIX B: REVIEW OF D_4 TOPOLOGICAL ORDER

The TO displayed by the Hamiltonian Eq. (21) is encoded in the twisted quantum double $\mathcal{D}^\alpha(\mathbb{Z}_2^3)$, which is equivalent to that of the quantum double of $\mathcal{D}(D_4)$. Here, we review first the anyon content of $\mathcal{D}(D_4)$ and then provide a mapping to the anyon content of $\mathcal{D}^\alpha(\mathbb{Z}_2^3)$, which is the convention used throughout the main text.

The group $D_4 \cong (\mathbb{Z}_2 \times \mathbb{Z}_2) \rtimes \mathbb{Z}_2$ corresponds to the symmetries of a square and is specified abstractly by elements r, s :

$$D_4 = \langle r, s | r^2 = s^4 = rsrs = 1 \rangle, \quad (\text{B1})$$

with center $\{s^2\}$. There are five conjugacy classes $\{1\}, \{s^2\}, \{s, s^3\}, \{r, rs^2\}, \{rs, rs^3\}$, and hence also five irreducible representations (irreps): four of dimension one and one of

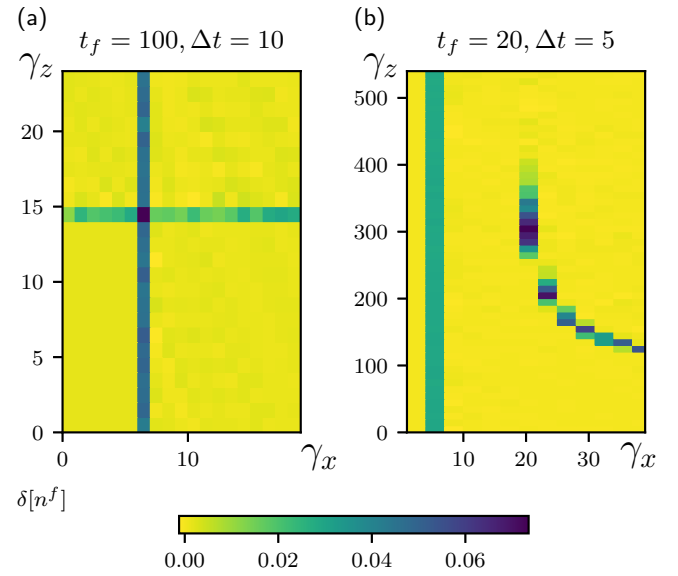


FIG. 18. The change in the total flag density $\delta[n^f]$ [see Eq. (A1)] over a time interval Δt is shown for (a) the toric code and (b) the D_4 model in the γ_x - γ_z parameter space at some late time t_f . In both plots, the change in the value of density after time Δt is small for parameters away from the phase boundaries. This suggests that the late time values evaluated at t_f (see the corresponding phase diagrams for flag densities in Fig. 1) are a good estimate of the steady-state densities. Near the transition, determination of the steady-state values requires more careful analysis of larger system sizes (presented for one specific value $\gamma_z = 500$ in Fig. 14 to identify the location of transition).

dimension two. We label the one dimensional irreps by ij ($i, j \in \mathbb{Z}_2$). In addition to the trivial irrep 00 and the faithful two-dimensional irrep σ , there are three sign irreps:

- (a) 10 is represented by $r = sr = s^2r = s^3r = -1$;
- (b) 01 is represented by $s = s^3 = sr = s^3r = -1$;
- (c) 11 is represented by $s = s^3 = r = s^2r = -1$.

The character table is given by

	1	s^2	s	r	rs
χ_{00}	1	1	1	1	1
χ_{01}	1	1	-1	1	-1
χ_{10}	1	1	1	-1	-1
χ_{11}	1	1	-1	-1	1
χ_σ	2	-2	0	0	0

(B2)

Note that the first four irreps obey the obvious $\mathbb{Z}_2 \times \mathbb{Z}_2$ fusion rules, whereas

$$ij \otimes \sigma = \sigma, \quad \sigma \otimes \sigma = \sum_{ij} ij, \quad (\text{B3})$$

obey a $\mathbb{Z}_2$ grading.

The anyon content of D_4 TO is specified as follows: each pointlike anyon can be labeled by the ordered pair $([g], \pi_g)$, where $[g]$ corresponds to a conjugacy class and

π_g denotes an irrep of its centralizer [93]. Thus, the four pure flux excitations are locally labeled by the nontrivial conjugacy classes, whereas the four pure charge excitations are labeled by the nontrivial elements of $\text{Rep}(D_4)$. The remaining 13 nontrivial excitations are dyons, leading to a total of 22 anyons (including the trivial anyon) in D_4 TO. The braiding phase between the pure charges and fluxes is given by the phase of the corresponding entry in the character table Eq. (B2): for instance, the wave function acquires a minus sign upon braiding the 11 charge around the $[s]$ flux or when the σ charge is braided with the $[s^2]$ flux.

The excitations of the Hamiltonian Eq. (21) considered in the main text can be labeled similarly to those of three copies of the two-dimensional toric code; these are generated by e_c and m_c , where the color index $c \in \{r, g, b\}$ indexes the different copies. Recall that the anyon content of the two-dimensional toric code is generated by the bosons e and m , which satisfy $\mathbb{Z}_2 \times \mathbb{Z}_2$ fusion rules. e and m have nontrivial mutual statistics since braiding one around the other generates an overall minus sign, and their fusion product $\psi = e \times m$ is an Abelian fermion. As discussed in Ref. [63], the 22 anyons of the twisted quantum double $\mathcal{D}^\alpha(\mathbb{Z}_2^3)$ are as follows.

- (1) There are three fundamental bosons $e_c (c = r, g, b)$, which are associated with the three colors of plaquette stabilizer defects in our model. Combinations of these also yield the three bicolored bosons $e_{rg} = e_r \times e_g$, etc., and one tricolored boson $e_{rgb} = e_r \times e_g \times e_b$. The eighth Abelian boson is the trivial particle ($e_c^3 = 1$ for all c).
- (2) Non-Abelian bosons:
 - (a) our three colors of vertex defects correspond to the three fundamental non-Abelian bosons m_r, m_g, m_b ; here, m_c and e_c are mutual semions, whereas m_c braids trivially with $e_{c'}$ ($c \neq c'$);
 - (b) combinations of adjacent vertex defects on different colored lattices yield an additional three non-Abelian bosons, m_{rg}, m_{gb}, m_{rb} ; $m_{cc'}$ (with $c \neq c'$) is a mutual semion with respect to e_c and $e_{c'}$, but has trivial mutual statistics with $e_{c''}$ ($c \neq c' \neq c''$).
- (3) Non-Abelian fermions:
 - (a) $\psi_c = m_c \times e_c$;
 - (b) $\psi_{cc'} = m_{cc'} \times e_c = m_{cc'} \times e_{c'}$ (for $c \neq c'$).
- (4) A non-Abelian semion s_{rgb} which braids non-trivially with e_c for all $c \in \{r, g, b\}$.
- (5) A non-Abelian antisemion $\bar{s}_{rgb} = s_{rgb} \times e_{rgb}$.

Each of the non-Abelian anyons in the theory has quantum dimension $d_\alpha = 2$. Ref. [63] provided an explicit mapping

TABLE I. Mapping between anyons in D_4 TO and anyon labels used in the main text, which correspond to anyons in twisted $\mathbb{Z}_2^3$ gauge theory.

$\mathcal{D}(D_4)$			$\mathcal{D}^\alpha(\mathbb{Z}_2^3)$	Dimension
Conjugacy class	Centralizer	Irrep		
1	D_4	00	1	1
1	D_4	10	e_{rg}	1
1	D_4	01	e_r	1
1	D_4	11	e_g	1
1	D_4	σ	m_b	2
$[s^2]$	D_4	00	e_{rgb}	1
$[s^2]$	D_4	10	e_b	1
$[s^2]$	D_4	01	e_{gb}	1
$[s^2]$	D_4	11	e_{rb}	1
$[s^2]$	D_4	σ	ψ_b	2
$[s]$	$\mathbb{Z}_4$	1	m_{rg}	2
$[s]$	$\mathbb{Z}_4$	ω	s_{rgb}	2
$[s]$	$\mathbb{Z}_4$	ω^2	ψ_{rg}	2
$[s]$	$\mathbb{Z}_4$	$\bar{\omega}$	$\bar{s}_{rgb}$	2
$[r]$	$\mathbb{Z}_2^2$	1	m_{gb}	2
$[r]$	$\mathbb{Z}_2^2$	$(-1, 1)$	m_g	2
$[r]$	$\mathbb{Z}_2^2$	$(1, -1)$	ψ_g	2
$[r]$	$\mathbb{Z}_2^2$	$(-1, -1)$	ψ_{gb}	2
$[sr]$	$\mathbb{Z}_2^2$	1	m_{rb}	2
$[sr]$	$\mathbb{Z}_2^2$	$(-1, 1)$	m_r	2
$[sr]$	$\mathbb{Z}_2^2$	$(1, -1)$	ψ_r	2
$[sr]$	$\mathbb{Z}_2^2$	$(-1, -1)$	ψ_{rb}	2

between the anyon theory of the quantum double $\mathcal{D}(D_4)$ and the anyon content of twisted $\mathbb{Z}_2$ gauge theory; we reproduce it here in Table I for completeness.

APPENDIX C: SIMULATION USING NON-PAULI STABILIZER TABLEAU

In this appendix, we provide details for simulating the Lindbladian dynamics of the D_4 model defined in terms of non-Pauli stabilizers [see Eq. (21)]. For an N -qubit system, the time evolution of standard Pauli stabilizers (and, consequently, the states that they stabilize) can be simulated by tracking the time evolution of the $O(N)$ stabilizers, each of which is a local product of Pauli operators. Since dynamics generated by Clifford gates maps Pauli stabilizers to Pauli stabilizers, any Clifford-based dynamics can be simulated efficiently by simply updating the stabilizers at each time step [95]. The ground space of the D_4 model is instead stabilized by the non-Pauli stabilizers (or “quasi-stabilizers”) defined in Eqs. (19) and (20): each of these is a tensor product of 1 and 2 qubit operators that are localized around a single vertex or plaquette of the honeycomb lattice. In an abuse of terminology, we refer to these operators simply as stabilizers, but we stress that these are not

mutually commuting elements of the N -qubit Pauli group. We now provide a generalized protocol that shows how the evolution of these stabilizers under the Pauli operations present in our dynamical protocol can be tracked similarly to genuine stabilizers, allowing us to efficiently simulate the noise and correction Lindbladian dynamics of the D_4 model.

We begin by choosing a basis for the independent stabilizers of the initial D_4 state. Any ground state of the D_4 Hamiltonian Eq. (21) is fully fixed by a list of plaquette and vertex stabilizer generators, together with the nonlocal logical operators that stabilize it. Each vertex of the honeycomb lattice has a vertex stabilizer B_v [see Eq. (19)]. While there are $2L^2/3$ operators on each sublattice of a system with $L \times L$ lattice sites, for each color c , these obey a global constraint $\prod_{v_c} B_{v_c} = \mathbb{I}$. Hence, the set of independent generators can be represented by

$$\mathcal{Q} = \{q_1 = B_{v_1}, q_2 = B_{v_2}, \dots, q_{2L^2-3} = B_{v_{2L^2-3}}\}, \quad (\text{C1})$$

with the values of the remaining three vertex stabilizers fully determined by this linearly independent set.

Now, while the plaquette stabilizers do not satisfy a similar constraint on the full Hilbert space, i.e., $\prod_{p_c} A_{p_c} \neq \mathbb{I}$, they are however constrained within the vertex-defect free subspace (i.e., $B_{v_c} \equiv 1$) via the relation:

$$\prod_{p_c} A_{p_c} = (-1)^{\frac{(1-\mathcal{Z}_{H'}^{c'})}{2} \frac{(1-\mathcal{Z}_{V'}^{c'')}{2}} \times (-1)^{\frac{(1-\mathcal{Z}_{V'}^{c'})}{2} \frac{(1-\mathcal{Z}_{H'}^{c'')}{2}}, \quad (\text{C2})$$

where the product is taken over all plaquette operators of the same color c and $\mathcal{Z}_{H/V}^{c'}$ represents the logical- Z operators of the remaining colors $c' \neq c'' \neq c$ in the horizontal (vertical) direction. Since the ground space is, by definition, defect-free, the set of independent plaquette generators relevant to our initial state can be represented as

$$\mathcal{G} = \{g_1 = A_1, g_2 = A_2, \dots, g_{L^2-3} = A_{L^2-3}\}, \quad (\text{C3})$$

with the values of the remaining plaquette stabilizers, denoted $A_{p_c}^*$ (one for each c), fixed by Eq. (C2).

Although Eq. (C2) holds true for a D_4 ground state, it will not be valid in the time-evolved state in general since this can have vertex stabilizer violations, generated by the noise. Instead, the relation will be modified as a result of the stabilizer updates. In principle, we can infer the modified relation at each time step from the set of instantaneous independent stabilizers. However, implementing this in practice requires decomposing each plaquette operator A_p in terms of its constituent CZ gates to determine their global pattern [as was done to derive Eq. (C2)]: this defeats the goal of treating operator strings as tensor products of the A_p operators. To overcome this, we instead explicitly

track the time evolution of three product operators $\prod_c A_{p_c}$, one for each color c . Each of these are length- $L^2/3$ operator strings composed of products of local A_{p_c} operators, which allows their evolution to be carried out efficiently using the methods described in the following.

Finally, depending on the specific logical state chosen in the simulation [see Eq. (41)], we choose one of the following two sets of initial logical stabilizers:

$$\mathcal{L} = \begin{cases} \{\mathcal{Z}_V^r, \mathcal{Z}_V^g, \mathcal{Z}_V^b, \mathcal{Z}_H^r, \mathcal{Z}_H^g, \mathcal{Z}_H^b\}, & \text{if } |\psi_0\rangle = |\tilde{0}\rangle_L, \\ \{\mathcal{Z}_V^r, \mathcal{Z}_V^g, \mathcal{Z}_V^b, \mathcal{X}_V^r, \mathcal{X}_V^g, \mathcal{X}_V^b\}, & \text{if } |\psi_0\rangle = |\tilde{+}\rangle_L. \end{cases} \quad (\text{C4})$$

In total, the $3L^2$ operators from the combined sets $\mathcal{Q}$, $\mathcal{G}$, and $\mathcal{L}$ describe a unique state as their $+1$ eigenvector [63].

To efficiently simulate the measurement dynamics, we also need to define a plaquette “destabilizer” generator d_k corresponding to each g_k [96]. These operators are chosen such that they satisfy the following algebraic relations:

$$[d_j, d_k] = [d_j, g_k] = 0 \quad \text{for } j \neq k, \quad \text{and} \quad \{d_j, g_j\} = 0 \quad (\text{C5})$$

for all $g_j \in \mathcal{G}$. Specifically, each destabilizer only anti-commutes with its partner stabilizer; this allows efficient detection of the presence or absence of g in operator strings, as we discuss later. The destabilizer d_{p_c} for the D_4 initial state can be defined as

$$d_{p_c} = \prod_{k \in l_p} Z_k, \quad (\text{C6})$$

where l_p represents a path on the dual lattice connecting the center of plaquette p_c to the designated plaquette p_c^* of the same color (recall that $A_{p_c}^* \notin \mathcal{G}$). It is straightforward to verify that this choice satisfies Eq. (C5) since every generator along the path l_p gets acted on by two Z operators, and hence they all commute with d_{p_c} . Only the operators A_{p_c} and $A_{p_c}^*$, located at the string’s endpoints, anticommute with d_{p_c} ; since $A_{p_c}^* \notin \mathcal{G}$, the destabilizer is uniquely associated with the plaquette stabilizer A_{p_c} in the initial state.

1. Evolution under X and Z operators

To simulate the dynamics, we track individual trajectories of our system over time. If S stabilizes the state $|\psi(t)\rangle$ at time t , then the state $|\psi(t+1)\rangle := U|\psi(t)\rangle$ is stabilized by the modified stabilizer operator $S' := USU^{-1}$. This unitary evolution also affects the destabilizer operators, which also need to be updated according to $D' := UDU^{-1}$. This simultaneous transformation leaves the defining properties of destabilizers in Eq. (C5) invariant. In the case at hand, both the noise and feedback operations involve the application of Pauli X and Z operators. This leads to the modified B_v and A_p operators given in Eqs. (37)–(39).

The logical stabilizers in $\mathcal{L}$ can be updated using similar transformations. The logical- Z operator is unchanged after applying Z operators, and transforms as

$$\pm \mathcal{Z} \xrightarrow{X_k} \mp \mathcal{Z} \quad (\text{C7})$$

if X_k acts on the edge that lies on the path defining $\mathcal{Z}$. The logical- X operator has both X and CZ operators in it. The corresponding transformation rule can be obtained using Eq. (25) as

$$\pm \mathcal{X} \xrightarrow{Z_k} \mp \mathcal{X}, \quad \text{and} \quad \pm \mathcal{X} \xrightarrow{X_j} \mp \prod_{i \in \xi(j)} Z_i \mathcal{X}. \quad (\text{C8})$$

Here Z_k (X_j) acts on one of the edges that have Pauli X (Z) in the definition of logical X , and $\xi(j)$ is the set of edges coupled by CZ to edge j in the definition of $\mathcal{X}$.

A key takeaway from this step is that, at any time t , the transformed stabilizer generators $g \in \mathcal{G}$ and $q \in \mathcal{Q}$ can be written as

$$g = \mathbf{s}_g \prod_k Z_k A_p, \quad \text{and} \quad q_v = \mathbf{s}_v B_v \quad (\text{C9})$$

where $\mathbf{s}_g, \mathbf{s}_q = \pm 1$. These are still tensor products of a set of 1- and 2-qubit operators localized near plaquettes and vertices.

2. Measurement of B_v and A_p

We now discuss the implementation of measurement operations. This requires determining the measurement outcome and the stabilizers of the postmeasurement state.

The vertex operators B_v commute with all stabilizers and, hence, their measurement leaves the state unchanged. The corresponding measurement outcome is readily available from the sign of elements in $\mathcal{Q}$. Similarly, if a measurement operator A_p is present in the list of instantaneous generators (i.e., $\pm A_p \in \mathcal{G}$), then the outcome is given by the sign of the stabilizer and the state is not modified.

If $\pm A_p$ is not amongst the list of generators, then we need to observe its algebraic relation with the generators in order to determine the result of the measurement on the state. In particular, note that $A_{p'}$ commutes with A_p up to a product of vertex operators [see e.g. Equation (23)]. It follows that the instantaneous stabilizer generators $g = \prod_k Z_k \prod_{p'} A_{p'}$ (that is, potentially made up of multiple A type operators) obey the relation:

$$\begin{aligned} A_p g &= A_p \left(\prod_k Z_k \prod_{p'} A_{p'} \right) = \left(\prod_k Z_k A_{p'} \right) A_p \left(\mathbf{s} \prod_{v \in \mathcal{V}} B_v \right) \\ &= g A_p \left(\mathbf{s} \prod_{v \in \mathcal{V}} B_v \right), \end{aligned} \quad (\text{C10})$$

where the sign $\mathbf{s} = \pm 1$ and the list of vertex operators $v \in \mathcal{V}$ obtained by commuting A_p with plaquette operators in g can be determined using Eqs. (20) and (23). The probability of getting a measurement outcome $\mathbf{m} = \pm 1$ upon measuring A_p is then given by

$$\begin{aligned} p_{\mathbf{m}} &:= \langle \psi | \frac{1 + \mathbf{m} A_p}{2} | \psi \rangle = \langle \psi | g_j \frac{1 + \mathbf{m} A_p}{2} g_j | \psi \rangle \\ &= \frac{1}{2} + \frac{\mathbf{m}}{2} \langle \psi | A_p \left(\mathbf{s} \prod_v B_v \right) | \psi \rangle, \end{aligned} \quad (\text{C11})$$

where we have used the fact that the state before measurement is stabilized by the operator $g_j \in \mathcal{G}$, with $(g_j)^2 = 1$.

a. Case (a)

If there exists at least one stabilizer $g_j \in \mathcal{G}$ such that $\mathbf{s} \prod_{v \in \mathcal{V}} B_v | \psi \rangle = -| \psi \rangle$, then we get

$$p_{\mathbf{m}} = \frac{1}{2} - \frac{\mathbf{m}}{2} \langle \psi | A_p | \psi \rangle = p_{-\mathbf{m}}, \quad (\text{C12})$$

where, in the last equality, we have used the fact that $p_{-\mathbf{m}} := \langle \psi | (1 - \mathbf{m} A_p)/2 | \psi \rangle$. Since A_p is a Hermitian operator that squares to 1, $\mathbf{m} = \pm 1$ are the only possible measurement outcomes which, together with Eq. (C12), implies that $p_{+1} = p_{-1} = 1/2$. Hence, the measurement outcome in this case is equally likely to be $+1$ or -1 , which we simulate by sampling a binary random number from the uniform distribution. Suppose $g_1, g_2, \dots, g_{n_0}$ are the $n_0 \geq 1$ stabilizers for which this holds. Then, they no longer stabilize the postmeasurement state stabilized by $\mathbf{m} A_p$ since they obey $g_j (1 + \mathbf{m} A_p) | \psi \rangle = (1 - \mathbf{m} A_p) | \psi \rangle$. Instead, the new state of the system is stabilized by pairwise products of these generators. This leads to a modification of stabilizers as follows: we choose a particular stabilizer g_1 from this list and the remaining anticommuting operators are updated as

$$g_j \rightarrow g_1 g_j \quad \text{for } j = 2, 3, \dots, n_0. \quad (\text{C13})$$

In addition, if any of the destabilizers anticommute with the new stabilizer $\mathbf{m} A_p$, then they also have to be modified to maintain the defining relations in Eq. (C5). This is achieved by updating

$$d_k \rightarrow g_1 d_k \quad \text{if } d_k A_p | \psi \rangle = -A_p d_k | \psi \rangle. \quad (\text{C14})$$

At the end of this step, stabilizer g_1 is replaced based on the measurement outcome as

$$d_1 \rightarrow g_1, \quad g_1 \rightarrow \mathbf{m} A_p. \quad (\text{C15})$$

b. Case (b)

If we find that $\mathbf{s} \prod_{v \in \mathcal{E}'} B_v |\psi\rangle = |\psi\rangle$ for all possible choices of $g_{p'}$ in Eq. (C11), then the set of stabilizers remains unchanged. This suggests that the state $|\psi\rangle$ must be an eigenstate of the measurement operator A_p with eigenvalue $\mathbf{m} = \pm 1$. Let us assume that the operator A_p can be expressed in terms of the instantaneous stabilizer generators as

$$\mathbf{m} A_p = \prod_i g_i^{\mu_i} \prod_j q_j^{\nu_j} \prod_k Z_k^{\xi_k}, \quad (\text{C16})$$

where the variables $\mu, \nu, \xi \in \{0, 1\}$ denote the presence or absence of the corresponding generator in the operator string. Recall from the definitions in Eqs. (C1) and (C3) that the individual operators q and Z are products of Pauli- Z operators whereas the operators g additionally involve plaquette stabilizers A . The goal is to find out the value of $\mathbf{m}$ by determining which generators multiply together to give $\mathbf{m} A_p$.

Since the plaquette destabilizer d_i anticommutes only with its partner g_i and commutes with the rest of $g \neq g_i$, we conclude that $\mu_i = 1$ if and only if $\{A_p, d_i\} = 0$. In principle, we can also detect the presence of vertex stabilizers q by tracking the evolution of the corresponding vertex destabilizers. Unfortunately, the operators that anticommute with only a pair of B_v would in general be linear-depth circuits since they correspond to creating pointlike non-Abelian anyons. Instead, we leverage the underlying lattice structure of the stabilizers to find the nonzero values of ν and ξ .

Let us consider the product $A_p \prod_{i: \mu_i=1} g_i$ which, according to Eq. (C16), must be of the form $\mathbf{s}' \prod_{l \in \mathcal{E}} Z_l$, where $\mathbf{s}' = \pm 1$ is an overall sign. For simplicity, let us first consider the case where $\xi_k = 0$, for which we do not require noncontractible loop operators to determine the measurement outcome. Then we have

$$\begin{aligned} \mathbf{m} &= \left(A_p \prod_{i: \mu_i=1} g_i \right) \prod_j (\mathbf{s}_{q_j} B_j)^{\nu_j} \\ &= \left(\mathbf{s}' \prod_{l \in \mathcal{E}} Z_l \right) \prod_j (\mathbf{s}_{q_j} B_j)^{\nu_j} = \mathbf{s}' \times \prod_{j: \nu_j=1} \mathbf{s}_{q_j}, \end{aligned} \quad (\text{C17})$$

where we have explicitly separated the signs $\mathbf{s}_{q_j}$ in the instantaneous vertex stabilizers from their operator B , as per the definition in Eq. (C9). The last equality can hold only if $Z_{l \in \mathcal{E}}$ precisely cancel the B operators for which ν is nonzero. This is possible if and only if $\nu_j = 1$ for vertices that are bounded by edges in $\mathcal{E}$ and zero otherwise. Thus, we find that the measurement outcome $\mathbf{m}$ is a product of the sign $\mathbf{s}'$ and the parity of the number of vertex defects bounded by edges in $\mathcal{E}$. This parity can be

efficiently computed by counting the number of times the path that pairs up all vertex defects (generated using the PYMATCHING package [106]) crosses the boundary $\mathcal{E}$. The case with a noncontractible loop, i.e., $\xi_k = 1$, can be treated analogously. Here, the relevant quantity to consider is the parity of defects that is bounded by $\mathcal{E}$ and the path of the logical- Z operator that is parallel to the noncontractible loop present in $\mathcal{E}$.

The updates of logical operators as a result of the measurements are carried out similarly. The logical- Z operators commute with both B_v and A_p operators and hence remain unchanged. The logical- X operators have nontrivial commutation relations with A_p as a result of Eq. (25). If, at a certain time step, measurement operators A_p anticommute with the logical (i.e., $\{A_p, \mathcal{X}_L\} |\psi\rangle = 0$) in addition to some other stabilizers g , then we update the logical- X operator as discussed in case (a). If we find that the measurement *only* anticommutes with the logical- X stabilizer, then it would imply that the state is no longer stabilized by the logical- X operator after the measurement. We term it as a logical error since it collapses the logical- X eigenstate.

3. Stabilizer tableaux

Having identified the form of instantaneous stabilizers, we now present the method used in our simulations for encoding them efficiently. The series of steps discussed in the previous subsection only switches the signs $\mathbf{s}_v$ associated with the vertex stabilizers $q_v = \mathbf{s}_v B_v$ without modifying the form of the operator. Hence, the instantaneous state of the vertex stabilizers can be represented by a length $2L^2$ array whose elements keep track of these signs.

The remaining plaquette-type stabilizers in $\mathcal{G}$ are modified in a nontrivial manner, which includes multiple A operators with additional Pauli- Z decorations in the given stabilizer. More explicitly, a generator $g_j \in \mathcal{G}$ can be written in terms of its constituents as

$$g_j = \mathbf{s}_j \prod_{k=1}^{N_q} (Z_k)^{z_{j,k}} \prod_{p=1}^{N_p} (A_p)^{a_{j,p}}, \quad (\text{C18})$$

where $N_q = 3L^2$ and $N_p = L^2$ is the total number of qubits and plaquettes in the system, respectively. Here, the products run over all three colors r, g, b but we suppress the color index for brevity. The variables $z_{j,k}$ and $a_{j,k}$ are either 1 or 0 depending on whether the associated operator is present or absent respectively and $\mathbf{s}_j = \pm 1$ is the overall sign. As discussed in relation to Eq. (C3), there are $L^2 - 3 = N_p - 3$ independent generators. We collect all of these operator representations in a $(N_p - 3) \times (N_q +$

N_p)-dimensional matrix

$$\mathbf{G} = \left[\begin{array}{ccc|ccc} z_{1,1} & \dots & z_{1,N_q} & a_{1,1} & \dots & a_{1,N_p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_{N_p-3,1} & \dots & z_{N_p-3,N_q} & a_{N_p-3,1} & \dots & a_{N_p-3,N_p} \end{array} \right], \quad (\text{C19})$$

where each entry is either 0 or 1. The overall sign $\mathbf{s}_j$ is stored in a separate array of length $N_p - 3$. The state of destabilizers can be analogously tracked using matrix $\mathbf{D}$ where j^{th} row represents the instantaneous operator d_j .

A key feature of this representation is that the matrices $\mathbf{G}$ and $\mathbf{D}$ can be updated efficiently. The update after application of X or Z to a qubit has a nontrivial effect on the stabilizer only if the plaquette sharing that qubit is present in its operator string. Hence, such an update only requires scanning through two columns of the matrix to determine the modified stabilizers. The updates are done by adding new Z decorations or changing the sign, which involves updating only $O(1)$ entries per row. The implementation of stabilizer measurement requires determination of the commutation relation between A_p and a row of $\mathbf{G}$ matrix. The operator A_p has nontrivial commutation relations only with six nearby plaquettes according to Eq. (23). Hence, the checks similar to Eq. (C11) to determine whether we are in case (a) or case (b) discussed previously require scanning through six columns of the matrix per measurement event and has $O(L^2)$ operation cost. The simulation of deterministic measurement outcome [see case (b)] has an additional overhead of generating matching graph that pairs all defects. We determine these pairings using PYMATCHING package that solves this problem in $\text{poly}(L)$ time [106]. The part of the update rules that depend on algebraic relations between stabilizer operators A, B and Pauli operators are precomputed at the beginning of the simulation. This includes indices of qubits coupled to a given qubit by CZ gates and indices of vertex operators that reside in the interior of an overlapping plaquette. Since this is still $O(1)$ data per qubit/plaquette, it does not change the overall memory requirement to simulate the given system. The locations of CZ gates in the logical- X string are also precomputed, which incurs a similar cost ($O(L)$ for each of the three X -logicals).

APPENDIX D: MEAN-FIELD EQUATIONS FOR ERASURE DYNAMICS

In this appendix, we discuss a simple mean-field picture of the dynamics of flags in the D_4 model. We model the dynamics using a site-approximation where we neglect the correlations amongst flags located on different sites. In particular, multisite correlation functions $\text{Tr}(\rho n_j^f n_k^f)$ in state ρ are approximated by the product of their average values $\text{Tr}(\rho n^f)^2$. Here, we additionally assume the densities to

be translation invariant. The mean-field dynamics is then described in terms of the X flag density n_X^f and the Z flag density n_Z^f .

The density of X flags evolves in time according to

$$\frac{dn_X^f}{dt} = \eta(1 - n_X^f) - 2\gamma_x n_X^f (1 - n_X^f)^2. \quad (\text{D1})$$

Here, the first term corresponds to the addition of X flags at unoccupied sites after an erasure error. The second term models the contribution from the leaf-correction protocol, where an X flag on a given edge is removed if it is the only flagged edge incident on one of its vertices. We note that the loop moves in the main text were introduced to remove closed loops of flags that are otherwise unaffected by the leaf-correction operations. Since the mean-field approximation neglects any intersite correlations, it does not distinguish between leaf and loop configurations. Thus, the correction term in Eq. (D1) is sufficient to remove all flags in this approximation, and we neglect the contributions from loop moves to obtain a simplified description.

The dynamics of Z flags is coupled to that of X flags as they evolve according to

$$\begin{aligned} \frac{dn_Z^f}{dt} = & \eta(1 - n_Z^f) + 2\gamma_x n_X^f (1 - n_X^f)^2 (1 - n_Z^f) \\ & - 2\gamma_z n_Z^f (1 - n_Z^f)^5 (1 - n_X^f)^6. \end{aligned} \quad (\text{D2})$$

Here, the first term corresponds to the addition of a Z flag as a result of an erasure error and the second terms adds a new Z flag to an unoccupied site as a result of the X correction procedure. The leaf-correction procedure is accounted for by the last term, where the factor $(1 - n_X^f)^6$ restricts the correction to the Z flags that do not have any X flags in their vicinity.

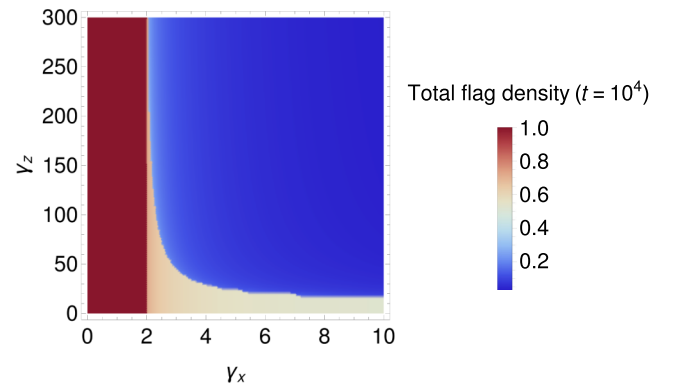


FIG. 19. Mean-field phase diagram of flag densities in the D_4 model: the color is proportional to the total density of flags $(n_X^f + n_Z^f)/2$ evaluated at a long time $t = 10^4$.

We numerically integrate the mean-field equations (D1) and (D2) up to some long time t to obtain the mean-field phase diagram shown in Fig. 19. At sufficiently high correction rates, the flags remain in an active state with small densities. In contrast, for small correction rates, they enter an absorbing state. These two phases are separated by an intermediate partially active phase where only X flags remain active. We note a qualitative difference compared to the genuine phase diagram, shown in Fig. 1(b): for large values of γ_z , the fully active phase of the mean-field model directly transitions into the absorbing phase without going into the intermediate partially active phase. This is not surprising since the partially active state in the unapproximated model appears because Z flags cannot be corrected as long as X flags inside the nearby plaquettes are not removed. However, the $(1 - n_X^f)^6$ factor in Eq. (D1) only approximates the averaged number of such plaquettes and hence the correction rate is overestimated by the mean-field equations.

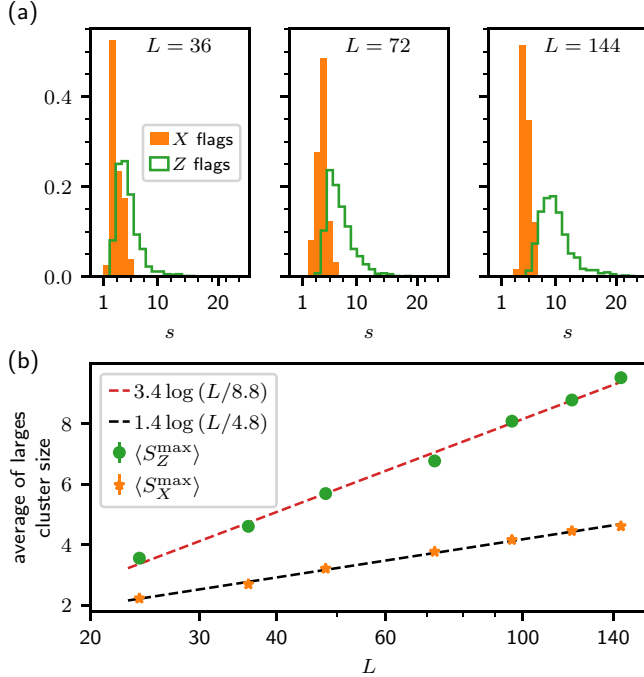


FIG. 20. Connected clusters of flags in active phase: the largest connected cluster of flags is determined in each Monte Carlo trajectory after 500 sweeps evaluated at a representative point in the active phase $\gamma_x = 35$, $\gamma_z = 500$. (a) Probability density $\rho(s)$ that the largest cluster in a given Monte Carlo realization is of size s is plotted for various system sizes for both X - and Z -type flags. The size s of a cluster is defined as the total number of flags in the connected cluster. Data are shown for 1000 samples. (b) The average size of the largest cluster $\langle S^{\max} \rangle$ obtained from the histograms is plotted as a function of increasing system size L . The error bars represent the standard error in the average value, and dashed lines are obtained by fitting the data points to function $a \log(L/L_0)$.

APPENDIX E: DISTRIBUTION OF CONNECTED FLAG CLUSTERS

In this appendix, we present further details of the mixed steady state ρ stabilized in the active phase of the D_4 model. When this state ρ is driven purely by the noiseless evolution, generated by the correction jump operators in $\mathcal{L}_C$ [see Eq. (35)], it approaches a ground state of the D_4 Hamiltonian Eq. (21), while all flags approach the $n^f = 0$ state. Specifically, as shown in Fig. 13, the time it takes for the infidelity $1 - F(\rho, |\psi_{D_4}\rangle)$ between a genuine D_4 ground state and the recovered state to fall below a fixed constant ε increases as $\sim \log L$ with increasing system size L . Note that the local moves of the correction channel remove the connected clusters of flags only from their endpoints. Intuitively, this suggests that the time for the recovery should be controlled by the largest cluster of flags in the original state ρ . In Fig. 20, we analyze the distribution of such largest connected clusters of flags in the Monte Carlo trajectories that sample the active steady state ρ . The average size of the largest cluster scales as $\log L$ which qualitatively agrees with the recovery timescale computed using the fidelity.

- [1] S. Bravyi, M. B. Hastings, and S. Michalakis, Topological quantum order: Stability under local perturbations, *J. Math. Phys.* **51**, 093512 (2010).
- [2] S. Bravyi and M. B. Hastings, A short proof of stability of topological order under local perturbations, *Commun. Math. Phys.* **307**, 609 (2011).
- [3] S. Michalakis and J. P. Zwolak, Stability of frustration-free Hamiltonians, *Commun. Math. Phys.* **322**, 277 (2013).
- [4] M. B. Hastings, Topological order at nonzero temperature, *Phys. Rev. Lett.* **107**, 210501 (2011).
- [5] S. Bravyi and B. Terhal, A no-go theorem for a two-dimensional self-correcting quantum memory based on stabilizer codes, *New J. Phys.* **11**, 043029 (2009).
- [6] O. Landon-Cardinal and D. Poulin, Local topological order inhibits thermal stability in 2D, *Phys. Rev. Lett.* **110**, 090502 (2013).
- [7] B. M. Terhal, Quantum error correction for quantum memories, *Rev. Mod. Phys.* **87**, 307 (2015).
- [8] B. J. Brown, D. Loss, J. K. Pachos, C. N. Self, and J. R. Wootton, Quantum memories at finite temperature, *Rev. Mod. Phys.* **88**, 045005 (2016).
- [9] S. Roberts, B. Yoshida, A. Kubica, and S. D. Bartlett, Symmetry-protected topological order at nonzero temperature, *Phys. Rev. A* **96**, 022306 (2017).
- [10] T.-C. Lu, T. H. Hsieh, and T. Grover, Detecting topological order at finite temperature using entanglement negativity, *Phys. Rev. Lett.* **125**, 116801 (2020).
- [11] C. Stahl and R. Nandkishore, Symmetry-protected self-correcting quantum memory in three space dimensions, *Phys. Rev. B* **103**, 235112 (2021).
- [12] This is fundamentally due to the presence of deconfined pointlike excitations which, when thermally excited,

- destroy the long-range entanglement in the ground space (even if the excitations have restricted mobility, as in fracton models [118–120]). However, certain models in $d \geq 4$ (e.g., the four-dimensional toric code [13,121]) lack any pointlike excitations and are thermally stable phases [8,122].
- [13] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
 - [14] A. Kay and R. Colbeck, Quantum self-correcting stabilizer codes, *ArXiv:0810.3557*.
 - [15] D. S. Wang, A. G. Fowler, A. M. Stephens, and L. C. L. Hollenberg, Threshold error rates for the toric and planar codes, *Quantum Info. Comput.* **10**, 456 (2010).
 - [16] J. R. Wootton, Quantum memories and error correction, *J. Mod. Opt.* **59**, 1717 (2012).
 - [17] J. R. Wootton, J. Burri, S. Iblisdir, and D. Loss, Error correction for non-abelian topological quantum computation, *Phys. Rev. X* **4**, 011051 (2014).
 - [18] C. G. Brell, S. Burton, G. Dauphinais, S. T. Flammia, and D. Poulin, Thermalization, error correction, and memory lifetime for Ising anyon systems, *Phys. Rev. X* **4**, 031058 (2014).
 - [19] J. R. Wootton and A. Hutter, Active error correction for Abelian and non-Abelian anyons, *Phys. Rev. A* **93**, 022318 (2016).
 - [20] G. Dauphinais and D. Poulin, Fault-tolerant quantum error correction for non-Abelian anyons, *Commun. Math. Phys.* **355**, 519 (2017).
 - [21] A. Schotte, L. Burgelman, and G. Zhu, Fault-tolerant error correction for a universal non-Abelian topological quantum computer at finite temperature, *ArXiv:2301.00054*.
 - [22] J.-M. Lihm, K. Noh, and U. R. Fischer, Implementation-independent sufficient condition of the Knill-Laflamme type for the autonomous protection of logical qudits by strong engineered dissipation, *Phys. Rev. A* **98**, 012317 (2018).
 - [23] A. Coser and D. Pérez-García, Classification of phases for mixed states via fast dissipative evolution, *Quantum* **3**, 174 (2019).
 - [24] C. de Groot, A. Turzillo, and N. Schuch, Symmetry protected topological order in open quantum systems, *Quantum* **6**, 856 (2022).
 - [25] A. Molnar, A. Ruiz de Alarcón, J. Garre-Rubio, N. Schuch, J. I. Cirac, and D. Pérez-García, Matrix product operator algebras I: Representations of weak Hopf algebras and projected entangled pair states, *ArXiv:2204.05940*.
 - [26] A. Ruiz-de-Alarcón, J. Garre-Rubio, A. Molnár, and D. Pérez-García, Matrix product operator algebras II: Phases of matter for 1D mixed states, *Lett. Math. Phys.* **114**, 43 (2024).
 - [27] J. Y. Lee, C.-M. Jian, and C. Xu, Quantum criticality under decoherence or weak measurement, *PRX Quantum* **4**, 030317 (2023).
 - [28] R. Ma and C. Wang, Average symmetry-protected topological phases, *Phys. Rev. X* **13**, 031016 (2023).
 - [29] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, Mixed-state topological order and the errorfield double formulation of decoherence-induced transitions, *ArXiv:2301.05687*.
 - [30] R. Fan, Y. Bao, E. Altman, and A. Vishwanath, Diagnostics of mixed-state topological order and breakdown of quantum memory, *PRX Quantum* **5**, 020343 (2024).
 - [31] J.-H. Zhang, Y. Qi, and Z. Bi, Strange correlation function for average symmetry-protected topological phases, *ArXiv:2210.17485*.
 - [32] J. Y. Lee, Y.-Z. You, and C. Xu, Symmetry protected topological phases under decoherence, *Quantum* **9**, 1607 (2025).
 - [33] Y.-H. Chen and T. Grover, Symmetry-enforced many-body separability transitions, *PRX Quantum* **5**, 030310 (2024).
 - [34] Y.-H. Chen and T. Grover, Separability transitions in topological states induced by local decoherence, *Phys. Rev. Lett.* **132**, 170602 (2024).
 - [35] Y. Guo and Y. Ashida, Two-dimensional symmetry-protected topological phases and transitions in open quantum systems, *Phys. Rev. B* **109**, 195420 (2024).
 - [36] D. Paszko, D. C. Rose, M. H. Szymańska, and A. Pal, Edge modes and symmetry-protected topological states in open quantum systems, *PRX Quantum* **5**, 030304 (2024).
 - [37] R. Sohal and A. Prem, Noisy approach to intrinsically mixed-state topological order, *PRX Quantum* **6**, 010313 (2025).
 - [38] T. D. Ellison and M. Cheng, Toward a classification of mixed-state topological orders in two dimensions, *PRX Quantum* **6**, 010315 (2025).
 - [39] Z. Li and R. S. K. Mong, Replica topological order in quantum mixed states and quantum error correction, *ArXiv:2402.09516*.
 - [40] S. Chirame, F. J. Burnell, S. Gopalakrishnan, and A. Prem, Stable symmetry-protected topological phases in systems with heralded noise, *Phys. Rev. Lett.* **134**, 010403 (2025).
 - [41] R. Ma and A. Turzillo, Symmetry-protected topological phases of mixed states in the doubled space, *PRX Quantum* **6**, 010348 (2025).
 - [42] P. Sala, S. Gopalakrishnan, M. Oshikawa, and Y. You, Spontaneous strong symmetry breaking in open systems: purification perspective, *Phys. Rev. B* **110**, 155150 (2024).
 - [43] T.-C. Lu, Disentangling transitions in topological order induced by boundary decoherence, *Phys. Rev. B* **110**, 125145 (2024).
 - [44] L. A. Lessa, R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, Strong-to-weak spontaneous symmetry breaking in mixed quantum states, *PRX Quantum* **6**, 010344 (2025).
 - [45] P. Sala, J. Alicea, and R. Verresen, Decoherence and wave-Function deformation of d_4 non-Abelian topological order, *Phys. Rev. X* **15**, 031002 (2025).
 - [46] P. Sala and R. Verresen, Stability and loop models from decohering non-Abelian topological order, *Phys. Rev. Lett.* **134**, 250403 (2025).
 - [47] J. Shah, C. Fechin, Y.-X. Wang, J. T. Iosue, J. D. Watson, Y.-Q. Wang, B. Ware, A. V. Gorshkov, and C.-J. Lin, Instability of steady-state mixed-state symmetry-protected topological order to strong-to-weak spontaneous symmetry breaking, *ArXiv:2410.12900*.
 - [48] R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, Topological phases with average symmetries: The decohered, the disordered, and the intrinsic, *Phys. Rev. X* **15**, 021062 (2025).

- [49] T. Rakovszky, S. Gopalakrishnan, and C. von Keyserlingk, Defining stable phases of open quantum systems, *Phys. Rev. X* **14**, 041031 (2024).
- [50] L. A. Lessa, M. Cheng, and C. Wang, Mixed-state quantum anomaly and multipartite entanglement, *Phys. Rev. X* **15**, 011069 (2025).
- [51] S. Sang, Y. Zou, and T. H. Hsieh, Mixed-state quantum phases: Renormalization and quantum error correction, *Phys. Rev. X* **14**, 031044 (2024).
- [52] S. Sang and T. H. Hsieh, Stability of mixed-state quantum phases via finite Markov length, *Phys. Rev. Lett.* **134**, 070403 (2025).
- [53] C. Zhang, Y. Xu, J.-H. Zhang, C. Xu, Z. Bi, and Z.-X. Luo, Strong-to-weak spontaneous breaking of 1-form symmetry and intrinsically mixed topological order, *Phys. Rev. B* **111**, 115137 (2025).
- [54] S. Sang, L. A. Lessa, R. S. K. Mong, T. Grover, C. Wang, and T. H. Hsieh, Mixed-state phases from local reversibility, *ArXiv:2507.02292*.
- [55] F. Thompson and A. Kamenev, Population dynamics of Schrödinger cats, *SciPost Phys.* **18**, 046 (2025).
- [56] M. Grassl, T. Beth, and T. Pellizzari, Codes for the quantum erasure channel, *Phys. Rev. A* **56**, 33 (1997).
- [57] E. Knill, Scalable quantum computing in the presence of large detected-error rates, *Phys. Rev. A* **71**, 042322 (2005).
- [58] Y. Wu, S. Kolkowitz, S. Puri, and J. D. Thompson, Erasure conversion for fault-tolerant quantum computing in alkaline earth Rydberg atom arrays, *Nat. Commun.* **13**, 4657 (2022).
- [59] A. Kubica, A. Haim, Y. Vaknin, H. Levine, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the T_1 limit in superconducting circuits, *Phys. Rev. X* **13**, 041022 (2023).
- [60] K. Sahay, J. Jin, J. Claes, J. D. Thompson, and S. Puri, High-threshold codes for neutral-atom qubits with biased erasure errors, *Phys. Rev. X* **13**, 041013 (2023).
- [61] S. Gu, A. Retzker, and A. Kubica, Fault-tolerant quantum architectures based on erasure qubits, *Phys. Rev. Res.* **7**, 013249 (2025).
- [62] S. Gu, Y. Vaknin, A. Retzker, and A. Kubica, Optimizing quantum error correction protocols with erasure qubits, *ArXiv:2408.00829*.
- [63] M. Iqbal, N. Tantivasadakarn, R. Verresen, S. L. Campbell, J. M. Dreiling, C. Figgatt, J. P. Gaebler, J. Johansen, M. Mills, S. A. Moses, J. M. Pino, A. Ransford, M. Rowe, P. Siegfried, R. P. Stutz, M. Foss-Feig, A. Vishwanath, and H. Dreyer, Non-Abelian topological order and anyons on a trapped-ion processor, *Nature* **626**, 505 (2024).
- [64] S. Ma, G. Liu, P. Peng, B. Zhang, S. Jandura, J. Claes, A. P. Burgers, G. Pupillo, S. Puri, and J. D. Thompson, High-fidelity gates and mid-circuit erasure conversion in an atomic qubit, *Nature* **622**, 279 (2023).
- [65] P. Scholl, A. L. Shaw, R. B.-S. Tsai, R. Finkelstein, J. Choi, and M. Endres, Erasure conversion in a high-fidelity Rydberg quantum simulator, *Nature* **622**, 273 (2023).
- [66] W. C. Campbell, Certified quantum gates, *Phys. Rev. A* **102**, 022426 (2020).
- [67] M. Kang, W. C. Campbell, and K. R. Brown, Quantum error correction with metastable states of trapped ions using erasure conversion, *PRX Quantum* **4**, 020358 (2023).
- [68] J. D. Teoh, P. Winkel, H. K. Babla, B. J. Chapman, J. Claes, S. J. de Graaf, J. W. O. Garmon, W. D. Kalfus, Y. Lu, A. Maiti, K. Sahay, N. Thakur, T. Tsunoda, S. H. Xue, L. Frunzio, S. M. Girvin, S. Puri, and R. J. Schoelkopf, Dual-rail encoding with superconducting cavities, *Proc. Natl. Acad. Sci.* **120**, e2221736120 (2023).
- [69] K. S. Chou, *et al.*, A superconducting dual-rail cavity qubit with erasure-detected logical measurements, *Nat. Phys.* **20**, 1454 (2024).
- [70] A. Koottandavida, I. Tsioutsios, A. Kargioti, C. R. Smith, V. R. Joshi, W. Dai, J. D. Teoh, J. C. Curtis, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret, Erasure detection of a dual-rail qubit encoded in a double-post superconducting cavity, *Phys. Rev. Lett.* **132**, 180601 (2024).
- [71] H. Levine, *et al.*, Demonstrating a long-coherence dual-rail erasure qubit using tunable transmons, *Phys. Rev. X* **14**, 011051 (2024).
- [72] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys. (N. Y.)* **303**, 2 (2003).
- [73] T. M. Stace, S. D. Barrett, and A. C. Doherty, Thresholds for topological codes in the presence of loss, *Phys. Rev. Lett.* **102**, 200501 (2009).
- [74] N. Delfosse and G. Zémor, Linear-time maximum likelihood decoding of surface codes over the quantum erasure channel, *Phys. Rev. Res.* **2**, 033042 (2020).
- [75] N. Delfosse and N. H. Nickerson, Almost-linear time decoding algorithm for topological codes, *Quantum* **5**, 595 (2021).
- [76] In fact, it is sufficient for the steady states of $\mathcal{L}, \mathcal{L}'$ to be connected by a short-time evolution involving a time-dependent Lindbladian that interpolates between the two.
- [77] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-Equilibrium Phase Transitions*, Theoretical and Mathematical Physics (Springer, Netherlands, 2008).
- [78] H. Hinrichsen, Non-equilibrium critical phenomena and phase transitions into absorbing states, *Adv. Phys.* **49**, 815 (2000).
- [79] These phases are characterized by the $\mathbb{Z}_2^{(0)}$ anyon theory and, in the parlance of Ref. [37], constitute an intrinsically mixed TO.
- [80] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Non-Abelian anyons and topological quantum computation, *Rev. Mod. Phys.* **80**, 1083 (2008).
- [81] S. X. Cui, S.-M. Hong, and Z. Wang, Universal quantum computation with weakly integral anyons, *Quantum Inf. Process.* **14**, 2687 (2015).
- [82] I. Cong, M. Cheng, and Z. Wang, Universal quantum computation with gapped boundaries, *Phys. Rev. Lett.* **119**, 170504 (2017).
- [83] Given a finite group G , its derived series is a set of normal subgroups N_j , derived inductively via $N_0 = G$ and $N_j = [N_{j-1}, N_{j-1}]$ ($j > 0$). Here, $[N, N]$ denotes the commutator subgroup of N . The derived length of G is defined as the smallest positive integer ℓ_G such that $N_{\ell_G} = \mathbb{Z}_1$. A solvable group is defined as one for which ℓ_G is finite.
- [84] T.-C. Lu, L. A. Lessa, I. H. Kim, and T. H. Hsieh, Measurement as a shortcut to long-range entangled quantum matter, *PRX Quantum* **3**, 040337 (2022).

- [85] N. Tantivasadakarn, A. Vishwanath, and R. Verresen, Hierarchy of topological order from finite-depth unitaries, measurement, and feedforward, *PRX Quantum* **4**, 020339 (2023).
- [86] S. Bravyi, I. Kim, A. Kliesch, and R. Koenig, Adaptive constant-depth circuits for manipulating non-Abelian anyons, *ArXiv:2205.01933*.
- [87] N. Tantivasadakarn, R. Verresen, and A. Vishwanath, Shortest route to non-Abelian topological order on a quantum processor, *Phys. Rev. Lett.* **131**, 060405 (2023).
- [88] T. I. Andersen, *et al.*, Google Quantum AI and Collaborators, Non-Abelian braiding of graph vertices in a superconducting processor, *Nature* **618**, 264 (2023).
- [89] S. Xu, *et al.*, Digital simulation of projective non-Abelian anyons with 68 superconducting qubits, *Chin. Phys. Lett.* **40**, 060301 (2023).
- [90] S. Xu, *et al.*, Non-Abelian braiding of Fibonacci anyons with a superconducting processor, *Nat. Phys.* **20**, 1469 (2024).
- [91] S. Goel, M. Reynolds, M. Girling, W. McCutcheon, S. Leedumrongwatthanakun, V. Srivastav, D. Jennings, M. Malik, and J. K. Pachos, Unveiling the non-Abelian statistics of $d(S_3)$ anyons using a classical photonic simulator, *Phys. Rev. Lett.* **132**, 110601 (2024).
- [92] B. Shi, Seeing topological entanglement through the information convex, *Phys. Rev. Res.* **1**, 033048 (2019).
- [93] R. Dijkgraaf and E. Witten, Topological gauge theories and group cohomology, *Commun. Math. Phys.* **129**, 393 (1990).
- [94] C. Galindo, E. C. Rowell, and Z. Wang, Acyclic anyon models, thermal anyon error corrections, and braiding universality, *ArXiv:1801.04296*.
- [95] D. Gottesman, The Heisenberg representation of quantum computers, *ArXiv:quant-ph/9807006*.
- [96] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* **70**, 052328 (2004).
- [97] Here, we are implicitly assuming that errors in the measurement-feedback routine have a qualitatively similar effect to applying a single-site Pauli error of the type analyzed in our model immediately following the measurement-feedback circuit.
- [98] K. J. Satzinger, *et al.*, Realizing topologically ordered states on a quantum processor, *Science* **374**, 1237 (2021).
- [99] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletić, and M. D. Lukin, Probing topological spin liquids on a programmable quantum simulator, *Science* **374**, 1242 (2021).
- [100] M. Foss-Feig, A. Tikku, T.-C. Lu, K. Mayer, M. Iqbal, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, A. Hankin, N. Hewitt, C. V. Horst, M. Matheny, T. Mengle, B. Neyenhuis, H. Dreyer, D. Hayes, T. H. Hsieh, and I. H. Kim, Experimental demonstration of the advantage of adaptive quantum circuits, *ArXiv:2302.03029*.
- [101] M. Iqbal, N. Tantivasadakarn, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, A. Hankin, N. Hewitt, C. V. Horst, M. Matheny, *et al.*, Topological order from measurements and feed-forward on a trapped ion quantum computer, *Commun. Phys.* **7**, 205 (2024).
- [102] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [103] G. Lindblad, On the generators of quantum dynamical semigroups, *Commun. Math. Phys.* **48**, 119 (1976).
- [104] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of N-level systems, *J. Math. Phys.* **17**, 821 (1976).
- [105] A. L. Toom, Stable and attractive trajectories in multicomponent systems, *Multicomponent Random Syst.* **6**, 549 (1980).
- [106] O. Higgott and C. Gidney, Sparse Blossom: Correcting a million errors per core second with minimum-weight matching, *Quantum* **9**, 1600 (2025).
- [107] B. Yoshida, Topological phases with generalized global symmetries, *Phys. Rev. B* **93**, 155131 (2016).
- [108] Let $\{V_c\}$ be the set of vertices along the string C_b and $\{P_c\}$ be the set of plaquettes whose centers lie along C_b , where $c \in \{r, g\}$. Denote by C_{rg}^v the set of red and green links which are incident on vertices along C_b , i.e., $C_{rg}^v = \{\ell_c : \ell_c \in {}^*v_c(c = r, g)\}$ for $v_c \in \{V_c\}$. Denote by C_{rg}^p those red and green links which are shared by pairs of plaquettes along C_b , i.e., $C_{rg}^p = \{\ell_c : \ell_c \in \partial p_c \text{ and } \ell_c \in \partial p'_c(c = r, g)\}$ for $p_c \neq p'_c \in \{P_c\}$. Then $C_{rg} = C_{rg}^v \cap C_{rg}^p$.
- [109] Evidently, the intermediate plaquette operators commute with each other and can thus be measured simultaneously, as by assumption there are no blue vertex defects along the string.
- [110] G. Ódor, Universality classes in nonequilibrium lattice systems, *Rev. Mod. Phys.* **76**, 663 (2004).
- [111] K. Binder, Theory of first-order phase transitions, *Rep. Prog. Phys.* **50**, 783 (1987).
- [112] M. M. de Oliveira, M. G. E. da Luz, and C. E. Fiore, Generic finite size scaling for discontinuous nonequilibrium phase transitions into absorbing states, *Phys. Rev. E* **92**, 062126 (2015).
- [113] S. Lübeck, Tricritical directed percolation, *J. Stat. Phys.* **123**, 193 (2006).
- [114] P. Grassberger, Tricritical directed percolation in 2+ 1 dimensions, *J. Stat. Mech.: Theory Exp.* **2006**, P01004 (2006).
- [115] C. Mochon, Anyon computers with smaller groups, *Phys. Rev. A* **69**, 032306 (2004).
- [116] M. B. Hastings and X.-G. Wen, Quasiadiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance, *Phys. Rev. B* **72**, 045141 (2005).
- [117] X. Chen, Z.-C. Gu, and X.-G. Wen, Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order, *Phys. Rev. B* **82**, 155138 (2010).
- [118] C. Chamon, Quantum glassiness in strongly correlated clean systems: An example of topological overprotection, *Phys. Rev. Lett.* **94**, 040402 (2005).

- [119] I. H. Kim and J. Haah, Localization from superselection rules in translationally invariant systems, [Phys. Rev. Lett. **116**, 027202 \(2016\)](#).
- [120] A. Prem, J. Haah, and R. Nandkishore, Glassy quantum dynamics in translation invariant fracton models, [Phys. Rev. B **95**, 155133 \(2017\)](#).
- [121] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, On thermal stability of topological qubit in Kitaev's 4D model, [Open Syst. Inf. Dyn. **17**, 1 \(2010\)](#).
- [122] H. Bombin, R. W. Chhajlany, M. Horodecki, and M. A. Martin-Delgado, Self-correcting quantum computers, [New J. Phys. **15**, 055023 \(2013\)](#).