

Kinetics of sliding-window quantum error correction

Adithya Sriram,¹ Charles Stahl,¹ Aleksander Kubica,² and Yaodong Li^{1,3}

¹*Department of Physics, Stanford University, Stanford, CA 94305*

²*Yale Quantum Institute & Department of Applied Physics, Yale University, New Haven, CT 06520*

³*Department of Mechanical Engineering & Department of Physics, University of Washington, Seattle, WA 98195*
(Dated: August 11, 2026)

Practical implementations of quantum error correction (QEC) require rapid measurement and continuous processing of the syndrome information in order to prevent a backlog of unprocessed data. While “static” QEC is theoretically well understood via mappings to equilibrium statistical mechanics models, such an understanding of “real-time” QEC is currently lacking. Here, we study the kinetics of sliding window decoding (SWD), an implementation of real-time decoding that acts on temporally local windows of noisy syndrome information and commits to corrections irreversibly at a nonzero rate. We propose an effective description of SWD in terms of a stochastic kinetic process, where $\mathbb{Z}_2$ -charged point particles undergo parity-conserving reaction and diffusion. This model describes dynamics at length and time scales large compared to the window size W , whereas physics at scales smaller than W leads to nontrivial W -dependent scaling of the effective parameters. We identify the rate of decoding $1/W$ as a relevant perturbation to the decodable phase. We also show broad applicability of our results by changing many microscopic details of SWD without affecting the effective description.

Introduction — There has been tremendous progress in bringing quantum error correction (QEC) from a question of theoretical importance to practical implementation [1–11]. Protecting quantum information in complex logical circuits while maintaining fast logical clock speeds is vital as we transition into the fault-tolerant era. To analyze the properties of a code, one often imagines an idealized “static” setting in which the experimentalist records (possibly faulty) syndrome information for an extensive period of time; this full history is used to infer the location of errors and make decoding decisions. Most of the theoretical understanding of QEC is based on this idealized setting. For instance, the seminal work of Dennis et al. [12] relates such a problem to a disordered equilibrium spin model and its phase transitions, with general applicability to a large family of QEC codes [13–18].

In contrast, any practical QEC protocol will require some version of “real-time” decoding, in which one only has access to the syndrome information within a short time interval before a decoding decision must be made. For example, in a scenario in which logical operations must be performed on the quantum information at a nonzero rate, decoding speed becomes the bottleneck [19]. In a prototypical scheme of real-time QEC, known as sliding-window decoding (SWD) [12, 20–22], the decoder combines a new set of syndrome information with the previously stored one, commits to corrections, and then discards the old data. This way, the decoder only ever maintains a finite amount of syndrome information. Practical considerations behind real-time decoding schemes have motivated many technical improvements [20, 21, 23], but comparatively little attention has been devoted to developing their theoretical descriptions.

Two features of real-time QEC rule out the equilibrium statistical mechanics description of the static set-

ting. First, decoding decisions are sequential and irreversible, made at a finite rate. Second, errors compound over time through a nontrivial interplay of physical noise, measurement noise, and imperfect recovery, so that the accumulated error is not captured by any local noise model. Consequently, one may wonder whether real-time QEC can be related to any effective model, and whether such a model must be intrinsically non-equilibrium. Such a connection would not only provide a physical intuition and effective description of the dynamics of QEC, but also establish fundamental bounds on the performance of real-time QEC, providing a theoretical benchmark for recent numerical results [20, 21, 23–25].

In this work, we develop such a description for SWD for topological codes with deconfined point-like defects carrying $\mathbb{Z}_2$ charge. Through analytical arguments and numerical experiments, we show that SWD, at time and length scales larger than the window size W , admits an effective description in terms of a stochastic kinetic model, namely a parity-conserving reaction-diffusion process, see Eq. (1). The reaction-diffusion kinetics might be anticipated on symmetry grounds, and we show in the End Matter that indeed many microscopic details of SWD can be changed without modifying this universality class. On the other hand, the scaling relations between parameters of the reaction-diffusion process and W are nontrivial and are not determined by symmetry alone. We connect the microscopic SWD dynamics and the effective description by extracting such scaling relations, and by describing a crossover between the real-time regime (where $W = O(1)$) and the static regime (where W exceeds the code distance) with $1/W$ as a relevant parameter. Our work thus yields a greater theoretical understanding of the intrinsic tradeoffs one must grapple with in real-time decoding.

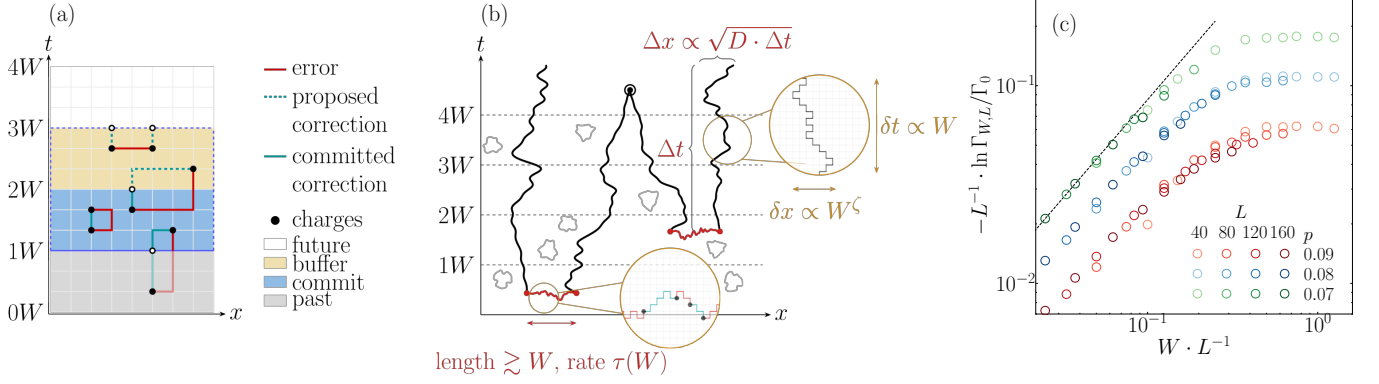


FIG. 1. (a) Illustration of the sliding-window decoder, for the first few rounds. The error and the correction together form closed loops in the bulk, with possible termination on the temporal boundary between the “buffer” and the “future” regions. Within each round, only correction operations in the “commit” region are turned into physical corrections; those in the “buffer” region serve as a look-ahead to improve decoding accuracy. (b) Worldlines of $\mathbb{Z}_2$ charges. *Fast* pair nucleation and annihilation events occur within $O(W)$ time scale (gray), whereas *slow* processes persist over many decoding rounds (black), and contribute to decoding failure. Nucleation events for slow charges are highlighted in red. (c) Data collapse for the memory time extracted from $\mathbb{P}_{\text{fail}}(t; W, L)$ against Eq. (5), see main text for details. Results at different values of $p < p_c$ fall onto the same universal function Φ (up to an overall rescaling of the vertical axis).

Definition of SWD — We first describe SWD in detail for a general d -dimensional CSS code of linear size L under phenomenological noise with physical and measurement errors of rate p . We will then focus on the 1D repetition code and the 2D toric code as prototypical examples in our numerical studies. For simplicity, we assume the decoder used is minimum weight perfect matching (MWPM) [12], though we also explore other decoder choices in the End Matter. Errors in the code create $\mathbb{Z}_2$ point-like charges and we will refer to them as such. In the Discussion, we will also comment on SWD as applied to more general classes of codes.

As we illustrate in Fig. 1(a), SWD proceeds via consecutive rounds of decoding, which we label with the integer $n \in [0, N]$. During the n -th round, we consider the syndrome information within the time window $[nW, (n+2)W]$, which reveals the locations of $\mathbb{Z}_2$ charges. The syndrome is input to the decoder, which matches the measured error chain endpoints, i.e. charges. Now, we divide this spacetime slice into a *commit window* and a *buffer window*, each of temporal size W [26]. The decoder only applies corrections that lie in the commit window. If the decoder matches charges that span the temporal boundary between the commit and buffer windows, a truncation of the matching to the commit window leaves behind a charge on this temporal boundary [hollowed dot in Fig. 1(a)], and this charge persists into the next round. The buffer window serves as a look-ahead that improves decoding accuracy, but no final decisions are made there yet. After this round, syndrome information in the commit window is erased. Each round thus overlaps with its neighboring ones and corrections are made every W time steps at a rate $1/W$. This process is depicted for the

(1+1)D repetition code in Fig. 1, and a formal definition can be found in the Supplemental Material [27].

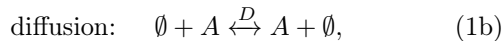
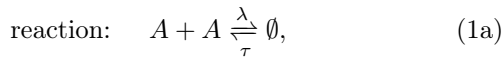
The combination of the errors and the proposed and committed corrections form closed loops in spacetime, except at the temporal boundary between the buffer window and the future window where open loops can anchor. These loops can be interpreted as the worldlines of the charges.

Effective model for SWD — In order to arrive at an effective description of the dynamics, we must make a crucial distinction between the *fast* $\mathbb{Z}_2$ charges which are created and annihilated within a single commit window, and the *slow* $\mathbb{Z}_2$ charges which actually contribute to logical failure. The worldlines of fast charges form closed loops of small spatiotemporal extent (see Fig. 1(a)) and they contribute a nonzero background density at any instant. However, they do not grow or persist and therefore do not seed logical failure.

On the other hand, the slow charges which do cause logical failure are those whose initial spatial distance reaches $\geq W$ within a single commit window. For a pair of such charges, the action of the decoder on these charges, if there does not exist a suitable matching involving other neighboring charges, transports each of them a bounded distance (of order W^ζ per round, see Eq. (3)) independent of L (Fig. 1(a)). The pair therefore can persist over many rounds of decoding, essentially performing independent random walks with finite step sizes. The time at which these pairs traverse the system and cause a logical error is what we call the memory time t_{mem} . We provide more details on their quasi-local motion in the Supplemental Material [27].

We argue that the motion of these slow charges is de-

scribed by the following reaction-diffusion process [28–30]



where A refers to a $\mathbb{Z}_2$ charge and $\emptyset$ refers to vacuum. Eq. (1a) describes a process in which two charges annihilate at rate λ to produce a vacuum state, as well as the reverse, in which two charges nucleate from a vacuum state at rate τ . Eq. (1b) describes the diffusion process, in which a charge can move to neighboring unoccupied sites at rate D . While pair creation is predominantly a consequence of the errors, diffusion and pair annihilation come from both the errors and the decoder.

We estimate the pair nucleation rate of the slow $\mathbb{Z}_2$ charges via a standard Peierls argument for MWPM, provided a sufficiently small p . In order for a pair to be created and persist beyond a single commit window, $O(W)$ “error events” are required to create and diffuse the charges a distance $\geq O(W)$ apart from each other. This leads to a pair nucleation rate $\tau \propto p^{\kappa W}$ at leading order, where κ is a nonuniversal geometrical factor that is fixed by the decoder. Generalizing to include higher order effects, we obtain the pair nucleation rate

$$\tau(W) \propto e^{-\kappa\sigma W}, \quad (2)$$

where $\sigma = \sigma(p)$ is the line tension that depends only on the error rate p but not W or L .

When two charges meet, they annihilate and cease to extend in the temporal direction. We therefore take the annihilation rate $\lambda(W) = 1$ independently of W , describing annihilation upon contact with probability 1.

We can also extract a diffusion constant by treating the motion of the coarse-grained charges as a random walk whose elementary moves are the worldlines within each window. These worldlines traverse a strip of temporal extent W . The microscopic processes within a window cause the worldline to acquire some spatial displacement $\delta\mathbf{r}$, whose statistics follow that of directed polymers in random media (DPRM), $|\delta\mathbf{r}| \propto W^\zeta$, where ζ is the wandering exponent [31]. The coarse-grained charge therefore executes a random walk with step size $|\delta\mathbf{r}| \propto W^\zeta$ taken every $\delta t \propto W$ time units, see Fig. 1(b). Within Δt time, the large-scale random walk takes $N = \Delta t / \delta t$ such steps, and travels a distance $|\Delta\mathbf{r}| \propto |\delta\mathbf{r}| \cdot N^{1/z}$, where $z = 2$ is the dynamic exponent of diffusion. We thus have an effective diffusion constant

$$D(W) \propto |\delta\mathbf{r}|^z \cdot (\delta t)^{-1} \propto W^{z\zeta-1}. \quad (3)$$

For SWD on the repetition code in (1+1)D we have the DPRM exponent $\zeta = 2/3$ [32–35].

Putting these processes together, we have a simple effective description of the dynamics in terms of a reaction-diffusion process of $\mathbb{Z}_2$ charges. The parameters D and τ

of the effective model scale with the window size W non-trivially as described in Eq. (3) and Eq. (2). We expect these to hold whenever $p < p_c$ [36]. We also expect that our discussions apply generally to codes with deconfined point-like excitations in an arbitrary spatial dimension.

The scaling of τ has direct consequences on the memory time of the code. With $\tau > 0$, charges nucleate and proliferate. At times longer than t_{relax} , the reaction-diffusion process reaches a steady state with a nonzero charge density. Similarly, at times longer than t_{mem} , the code enters a state where different logical sectors are maximally mixed. For the codes that we consider here, a finite density of charges leads to a nonvanishing logical error rate, so we intuitively identify the relaxation time t_{relax} of the process in Eq. (1) with the memory time t_{mem} of the code under the SWD dynamics [37]. In particular, a numerical calculation of the memory time also provides an indirect test of Eq. (2) via

$$t_{\text{mem}} \propto t_{\text{relax}} \propto \tau^{-\alpha} \propto e^{+\alpha\kappa\sigma W}, \quad (4)$$

where α is a universal exponent for the process in Eq. (1) [38], and we have $\alpha = 1$ in one spatial dimension. Below, we numerically test the exponential scaling of t_{mem} with W without directly extracting α [39].

Crossover to $W \gg L$ — Throughout the above discussion, we have assumed that $W \ll L$. If instead $W \rightarrow \infty$ we see from Eq. (3) that D diverges, indicating the breakdown of the kinetic description. The charges are no longer random walkers slowly diffusing across the system, as they can now make nonlocal jumps. Indeed, as W approaches and exceeds $O(L)$, the reaction-diffusion dynamics Eq. (1) must give way to the equilibrium statistical mechanics model. Here, the dominant logical error is no longer due to proliferation of $\mathbb{Z}_2$ charges, but instead a single space-like domain wall crossing the system, whose free energy cost scales linearly in L . This yields a memory time that diverges as $\ln t_{\text{mem}} \propto L$.

Our discussions above motivate the following scaling form describing the crossover in W ,

$$\ln t_{\text{mem}} \propto L \cdot \Phi(W/L), \quad (5)$$

for a *universal* scaling function Φ . To match the scaling of t_{mem} for $W \ll L$ and $W \gg L$, the limiting behaviors of $\Phi(x)$ are required to be (1) for $x \rightarrow 0$, $\Phi(x) \approx \alpha\kappa\sigma x$, so that $\ln t_{\text{mem}} \approx \alpha\kappa\sigma W$ in agreement with Eq. (4), and (2) for $x \rightarrow \infty$, $\Phi(x) \rightarrow \sigma$, so that $\ln t_{\text{mem}} \approx \sigma L$, in agreement with the expectation in the static limit.

Numerical results — We now provide numerical tests of SWD for the 1D repetition code. These results also provide evidence for the effective kinetic description from above. Parameters of this problem include the system size L , the window size W , the error rate p , and the time steps t , which is taken to be up to multiples of W . We will fix p to be below the threshold p_c (defined when $W \rightarrow \infty$) and vary the other parameters. For the repetition code,

we have $p_c \approx 0.103$. We also study the 2D toric code, with results reported in End Matter.

After t time steps (i.e. t/W rounds of SWD), we define the logical failure probability at this time, denoted $\mathbb{P}_{\text{fail}}(t; W, L)$, through a noiseless recovery channel (chosen to be MWPM without measurement errors) applied to the final state. For the repetition code, it is also useful to define $\Delta\rho := |\rho_0 - \rho_1|$, i.e. the absolute value of the *spin density* in the physical state. We initialize the system in the all-0 state, i.e. $\Delta\rho = 1$. We track both $\Delta\rho(t; W, L)$ and $\mathbb{P}_{\text{fail}}(t; W, L)$ over time. Our choice of these observables, which track the physical and logical errors in the system, is motivated by previous studies of single-shot QEC [40–43].

The behavior of $\mathbb{P}_{\text{fail}}(t; W, L)$ is characterized by a linear-in- t growth with slope denoted $\Gamma_{W,L}$, up until some time scale $\propto \Gamma_{W,L}^{-1}$, at which point $\mathbb{P}_{\text{fail}}(t; W, L)$ saturates [44]. We thus define $t_{\text{mem}} \equiv \Gamma_{W,L}^{-1}$. In Fig. 1(c), we numerically find that

$$-L^{-1} \cdot \ln(\Gamma_{W,L}/\Gamma_0) \propto \Phi(W/L), \quad (6)$$

in agreement with Eq. (5). The limiting behaviors of the functional form at small and large values of W/L also agree with our expectations below Eq. (5). The growth rate of $\Delta\rho(t; W, L)$ obeys a similar scaling collapse.

In performing the data collapse, the only fitting parameter is the overall multiplicative factor Γ_0 of $\Gamma_{W,L}$ contributing to an *additive* constant to the RHS of Eq. (5), which we fix independently using data from the static limit ($W \rightarrow \infty$). We expect the p -dependence of the failure probability to be captured entirely by the “line tension” $L^{-1} \cdot \ln(\Gamma_{W,L}/\Gamma_0)$, and Γ_0 to be a geometric factor independent of p . Therefore, for all three values of p in Fig. 1(c), we use the same fitting parameter Γ_0 .

We note that the form in Eq. (5) also allows for a non-universal *multiplicative* constant between the LHS and the RHS. Indeed, data at different values of p can be collapsed onto the same function Φ , after an overall rescaling of the vertical axis, confirming our expectation that Φ is universal [45]. Closer to the critical point $p = p_c$, the vanishing of the line tension should be universal, but the presence of multiple length scales makes a numerical calculation more demanding, and we leave a more thorough investigation of the critical regime for future work.

Next, we establish numerical evidence for the diffusive scaling of the dynamics and extract the diffusion constant, see Eq. (3). We initialize the system with a single domain wall at $x = L/2$ (so that $\Delta\rho = 0$), as illustrated in Fig. 2, and study its motion under SWD. For this experiment we take open boundary conditions.

We first focus on the regime $W \ll L^\zeta$ [Fig. 2(a)]. Here, from discussions leading to Eq. (3), the isolated domain wall appears as a random walker, with directed polymers of temporal extent $\propto W$ as its microscopic steps. The distance of the random walker from its initial location

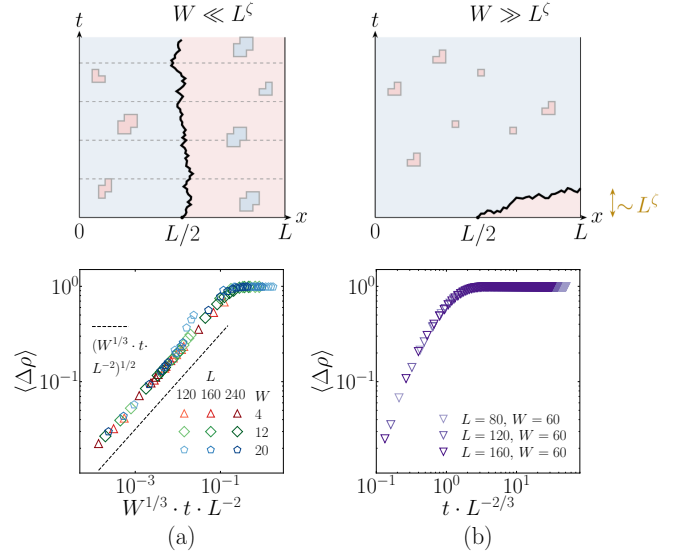


FIG. 2. Diffusive motion of $\mathbb{Z}_2$ charges. (a) For $W \ll L^\zeta$, we predict that $\langle \Delta\rho \rangle \propto (D \cdot t \cdot L^{-2})^{1/2}$, where $D \propto W^{1/3}$ is the effective diffusion constant, see Eq. (3). (b) For $W \gg L^\zeta$, the domain wall tends to take a shortest path to terminate either on the left or right boundary. Thus, $\langle \Delta\rho \rangle$ saturates at time $t \propto L^\zeta$, where L^ζ is the height of the domain wall. At earlier times, $\langle \Delta\rho \rangle$ depends on t through the scaling variable $t \cdot L^{-\zeta}$.

$x = L/2$ can be related to the spin density as

$$\langle |\Delta x| \rangle \propto L \cdot \langle \Delta\rho \rangle, \quad (7)$$

where $\langle \cdot \rangle$ is the expectation over all possible error configurations weighted by the phenomenological noise model. This observable has the advantage that fast pair creation-annihilation processes (i.e. bubbles whose size distribution is independent of W) do not contribute. For diffusive scaling $\langle |\Delta x| \rangle \propto (D(W) \cdot \Delta t)^{1/2}$, we expect that

$$\langle \Delta\rho \rangle \propto L^{-1} \langle |\Delta x| \rangle \propto L^{-1} (W^{\zeta-1} \cdot \Delta t)^{1/2}, \quad (8)$$

where we have used $D \propto W^{\zeta-1}$ (3). Our data is in excellent agreement with this expectation (see Fig. 2(a)).

In the opposite regime $W \gg L^\zeta$ (see Fig. 2(b)), the dominant contributions come from DPRMs directed in the spatial direction. Its temporal extent scales as L^ζ [27] and for $t \gg L^\zeta$ the random walker has a high probability of having terminated on either the left or right boundaries, with $\langle \Delta\rho \rangle$ saturating at its steady state value. Before saturation, the growth of $\langle \Delta\rho \rangle$ is described by a scaling function with a single scaling variable $t \cdot L^{-\zeta}$.

These two experiments both support the theoretical picture we outlined, including the reaction-diffusion Eq. (1) as well as the scaling of the pair nucleation rate Eq. (2), the diffusion constant Eq. (3), and the memory time Eq. (5).

Discussion — In this work, we have studied the dynamics of SWD, which is one type of real-time QEC,

focusing on topological codes with deconfined point-like defects. We have argued that the long-time, large-scale behavior admits a simple effective description, namely a parity-conserving reaction-diffusion process. The memory time of the code is the relaxation time of this effective process, and its dependence on W and L is captured by a single-variable crossover scaling that interpolates between the reaction-diffusion regime at $W \ll L$ and the static regime at $W \gg L$. We argued for this general description, and provided extensive numerical evidence for its validity, using the 1D repetition code and the 2D toric code (End Matter) as numerical examples.

Our results provide concrete theoretical guarantees for the design and evaluation of SWD-style real-time decoders. First, the crossover scaling identifies W/L as the natural figure of merit for speed-accuracy tradeoff, which can be used to inform error-budgeting with a given throughput constraint. Second, the reaction-diffusion picture is general and applies to a broad range of noise models and decoders, including those with explicit quasi-locality in both space and time or those with lower algorithmic complexity. We explore these possibilities in the End Matter and Supplemental Material [27], finding agreements with the reaction-diffusion process.

The dynamics of SWD is sensitive to the sizes of the commit and buffer windows, which control the creation of the slow $\mathbb{Z}_2$ charges whose subsequent diffusion across the system leads to logical failure. Previous numerical studies of SWD [20, 21, 23] found that logical error suppression is exponential in the code distance d and performance close to the static setting can be recovered for windows scaling linearly with d . Our analysis of SWD provides a dynamical explanation for this dependence; varying the window sizes allows to probe sensitivity to timelike measurement errors, closely connecting to stability experiments [46] and lattice-surgery simulations [47].

Our effective model provides a new perspective on the fact that SWD is not a single-shot decoder for codes with point-like excitations. For any finite W , the memory time saturates and does not diverge as a function of L . Indeed, from Eq. (5), the rate of decoding $1/W$ acts as a *relevant perturbation* to the decodable (ordered) phase, and similarities can be drawn with dimensional crossover [48], see also [49, 50]. Single-shot decoders, on the other hand, have diverging memory times even when the decoder only has access to syndrome information from a single time step [40–42, 51, 52]. Even though allowing for $W = O(1)$ time steps may improve the threshold [25], it does not qualitatively change the behavior of the decoder. An effective description of single-shot decoding would provide a universal understanding of this behavior. In particular, it would allow for the more general definition of single-shot memories as those for which $1/W$ is an irrelevant perturbation to the memory, unlike SWD where $1/W$ is relevant.

Looking forward, real-time QEC may be viewed as

perhaps the simplest non-trivial example of many-body dynamics driven by local operations and classical communication (LOCC). The resulting dynamics is neither unitary nor described by a fixed local quantum channel. What other kinds of many-body dynamics can be realized under LOCC, when the figure of merit are communication bandwidth, range, latency, in addition to the relaxation time? Which non-equilibrium universality classes, if any, are accessible to LOCC but not to local rules? While examples in this work belong to a known dynamical universality class, excitations carrying charge in a more general group than $\mathbb{Z}_2$ (either abelian or non-abelian), or excitations that obey kinetic constraints (instead of being deconfined) [53, 54], could lead to qualitatively new behavior.

Acknowledgments — We thank Arpit Dua, Hyunsoo Ha, Jacob Hauser, Vedika Khemani, Ethan Lake, Aditya Mahadevan, Nicholas O’Dea, and Shengqi Sang for helpful discussions. A.S. was supported in part by the Office of Naval Research Young Investigator Program (ONR YIP) under Award Number N00014-24-1-2098 and in part by a Packard Fellowship in Science and Engineering (PI: Vedika Khemani). A.S. also acknowledges support from the NSF graduate research fellowship and the ARCS Scholar award. C.S. was supported by the US Department of Energy, Office of Science under Award No DE-SC0025934. A.K. acknowledges support from the NSF (QLCI, Award No. OMA-2120757), IARPA and the Army Research Office (ELQ Program, Cooperative Agreement No. W911NF-23-2-0219). Y.L. was supported in part by the Gordon and Betty Moore Foundation’s EPiQS Initiative through Grant GBMF8686, in part by the US Department of Energy, Office of Science under Award No DE-SC0019380, and in part by a Q-FARM Bloch Postdoctoral Fellowship at Stanford University. We acknowledge the hospitality of the Kavli Institute for Theoretical Physics (KITP), which is supported in part by grant NSF PHY-2309135.

-
- [1] C. Ryan-Anderson *et al.*, *Science* **385**, 1327–1331 (2024).
 - [2] R. Acharya *et al.*, *Nature* **638**, 920–926 (2025).
 - [3] I. Pogorelov, F. Butt, L. Postler, C. D. Marciniak, P. Schindler, M. Müller, and T. Monz, *Nature Physics* **21**, 298–303 (2025).
 - [4] H. Putterman *et al.*, *Nature* **638**, 927–934 (2025).
 - [5] N. Lacroix *et al.*, *Nature* **645**, 614–619 (2025).
 - [6] P. Sales Rodriguez *et al.*, *Nature* **645**, 620–625 (2025).
 - [7] B. L. Brock, S. Singh, A. Eickbusch, V. V. Sivak, A. Z. Ding, L. Frunzio, S. M. Girvin, and M. H. Devoret, *Nature* **641**, 612–618 (2025).
 - [8] L. Daguerre, R. Blume-Kohout, N. C. Brown, D. Hayes, and I. H. Kim, *Physical Review X* **15**, 041008 (2025).
 - [9] A. Eickbusch *et al.*, *Nature Physics* **21**, 1994–2001 (2025).
 - [10] D. Bluvstein *et al.*, *Nature* **649**, 39–46 (2025).

- [11] B. Zhang, G. Liu, G. Bornet, S. P. Horvath, P. Peng, S. Ma, S. Huang, S. Puri, and J. D. Thompson, *Nature Physics* **22**, 910–916 (2026).
- [12] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Journal of Mathematical Physics* **43**, 4452 (2002).
- [13] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, *Physical Review Letters* **103**, 090501 (2009).
- [14] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, *Physical Review X* **2**, 021004 (2012).
- [15] I. Dumer, A. A. Kovalev, and L. P. Pryadko, *Phys. Rev. Lett.* **115**, 050502 (2015).
- [16] A. Kubica, M. E. Beverland, F. Brandão, J. Preskill, and K. M. Svore, *Phys. Rev. Lett.* **120**, 180501 (2018).
- [17] C. T. Chubb and S. T. Flammia, *Annales de l'Institut Henri Poincaré D, Combinatorics, Physics and their Interactions* **8**, 269–321 (2021).
- [18] Y. Li, N. O'Dea, and V. Khemani, Perturbative stability and error correction thresholds of quantum codes (2024), [arXiv:2406.15757 \[quant-ph\]](#).
- [19] B. M. Terhal, *Rev. Mod. Phys.* **87**, 307 (2015).
- [20] L. Skoric, D. E. Browne, K. M. Barnes, N. I. Gillespie, and E. T. Campbell, *Nature Communications* **14**, 7040 (2023).
- [21] X. Tan, F. Zhang, R. Chao, Y. Shi, and J. Chen, *PRX Quantum* **4**, 040344 (2023).
- [22] F. Battistel, C. Chamberland, K. Johar, R. W. J. Overwater, F. Sebastiano, L. Skoric, Y. Ueno, and M. Usman, *Nano Futures* **7**, 032003 (2023).
- [23] H. Bombin, C. Dawson, Y.-H. Liu, N. Nickerson, F. Pastawski, and S. Roberts, Modular decoding: parallelizable real-time decoding for quantum computers (2023), [arXiv:2303.04846 \[quant-ph\]](#).
- [24] T. Oberoi, J. Viszlai, and F. T. Chong, *Adapt: Adaptive-window decoding for practical fault-tolerance* (2026), [arXiv:2605.01149 \[quant-ph\]](#).
- [25] S. Huang and S. Puri, *Phys. Rev. A* **110**, 012453 (2024).
- [26] In principle, the relative size of the buffer window and commit window may be tuned, but for simplicity, we make them equal.
- [27] See Supplemental Material for further discussions of (i) formal definition and effective description of the sliding window decoder (SWD), and (ii) an absorbing state phase transition of the SWD under SWAP noises.
- [28] J. L. Cardy and U. C. Täuber, *Journal of Statistical Physics* **90**, 1 (1998), [arXiv:cond-mat/9704160 \[cond-mat.stat-mech\]](#).
- [29] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-equilibrium Phase Transitions: Volume I: Absorbing Phase Transitions* (Springer, 2008).
- [30] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, 2010).
- [31] In (1+1)D and for a large enough W , the worldlines here are domain walls within a 2D random-bond Ising model, and they have DPRM exponents [32–35].
- [32] D. A. Huse and C. L. Henley, *Phys. Rev. Lett.* **54**, 2708 (1985).
- [33] M. Kardar, *Phys. Rev. Lett.* **55**, 2923 (1985).
- [34] D. A. Huse, C. L. Henley, and D. S. Fisher, *Phys. Rev. Lett.* **55**, 2924 (1985).
- [35] M. Kardar, G. Parisi, and Y.-C. Zhang, *Phys. Rev. Lett.* **56**, 889 (1986).
- [36] Here, p_c is the threshold defined in the static limit ($W \rightarrow \infty$), as our arguments really only depend on the assumption of a nonzero domain wall line tension.
- [37] However, in single-shot codes, self-correcting codes, or other code families this identification may not hold.
- [38] Z. Rácz, *Phys. Rev. Lett.* **55**, 1707 (1985).
- [39] In our numerical experiments, slow charges are always clouded by a background of fast charges, and are not easily distinguishable.
- [40] H. Bombin, *Phys. Rev. X* **5**, 031043 (2015).
- [41] A. Kubica and M. Vasmer, *Nature Communications* **13**, 6272 (2022).
- [42] B. J. Brown, N. H. Nickerson, and D. E. Browne, *Nature Communications* **7**, 12302 (2016).
- [43] S. Gu, E. Tang, L. Caha, S. H. Choe, Z. He, and A. Kubica, *Communications in Mathematical Physics* **405**, 85 (2024).
- [44] The steady state value of $\mathbb{P}_{\text{fail}}(t; W, L)$ is simply $1 - 2^{-(\#\text{logicals})}$, which for the repetition code equals $1/2$.
- [45] We note that a rescaling of the horizontal axis is not required to collapse data at different p in Fig. 1, as long as the decoder is fixed. Comparing the asymptotics of the scaling function $\Phi(x)$ [see discussions below Eq. (5)], namely $\Phi(x \rightarrow \infty) = \sigma(p)$ and $\Phi'(x = 0) = \alpha\kappa\sigma(p)$, we see that the dependence on p comes entirely from the line tension, and can thus be eliminated by rescaling the vertical axis only. In particular, the “bend” of Φ is near $x_* \approx 1/(\alpha\kappa)$, independent of p .
- [46] C. Gidney, *Quantum* **6**, 786 (2022).
- [47] C. Chamberland and E. T. Campbell, *PRX Quantum* **3**, 010331 (2022).
- [48] J. Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge Lecture Notes in Physics (Cambridge University Press, 1996).
- [49] S. J. S. Tan, C. A. Pattison, M. McEwen, and J. Preskill, *Phys. Rev. A* **113**, 022450 (2026).
- [50] A. Sriram, N. O'Dea, Y. Li, T. Rakovszky, and V. Khemani, *Phys. Rev. B* **111**, 224208 (2025).
- [51] Y. Li, C. W. von Keyserlingk, G. Zhu, and T. Jochym-O'Connor, *Phys. Rev. Research* **6**, 043007 (2024).
- [52] C. Stahl, *Phys. Rev. B* **110**, 075143 (2024).
- [53] M. E. J. Newman and C. Moore, *Phys. Rev. E* **60**, 5068 (1999).
- [54] J. Haah, *Phys. Rev. A* **83**, 042330 (2011).
- [55] N. Delfosse and N. H. Nickerson, *Quantum* **5**, 595 (2021).
- [56] G. Duclos-Cianci and D. Poulin, *Phys. Rev. Lett.* **104**, 050504 (2010), [arXiv:0911.0581 \[quant-ph\]](#).
- [57] S. Bravyi and J. Haah, *Phys. Rev. Lett.* **111**, 200501 (2013).
- [58] C. Wang, J. Harrington, and J. Preskill, *Annals of Physics* **303**, 31 (2003).
- [59] H. Hinrichsen, *Advances in Physics* **49**, 815 (2000), [arXiv:cond-mat/0001070](#).
- [60] D. Zhong and D. ben Avraham, *Physics Letters A* **209**, 333 (1995).

END MATTER

Here, we explore a number of different scenarios for SWD beyond the repetition code and MWPM, with the aim of extending the validity of our effective description to a greater range of codes and decoders.

Explicitly quasi-local dynamics

The quasi-local effective description, particularly that point defects have a “range of interaction” of $O(W)$, suggests that it might be possible to capture large scale properties of SWD (which is nonlocal in space) by a microscopic model with quasi-locality of range $O(W)$. We examine a few such realizations and confirm these expectations. These numerical experiments serve as additional nontrivial tests of our general reaction-diffusion description. They also provide theoretical guarantees to these decoders (which are desirable from the practical standpoint due to their quasi-locality and numerical efficiency). In all cases, we focus on $\mathbb{P}_{\text{fail}}(t; W, L)$ as described in the main text.

Modular SWD

We first consider the modular SWD, defined in Fig. 3(a,b). We divide the spacetime lattice into square blocks of size $W \times W$. Within each round, we first commit to even-numbered blocks, with their $3W \times 2W$ sur-

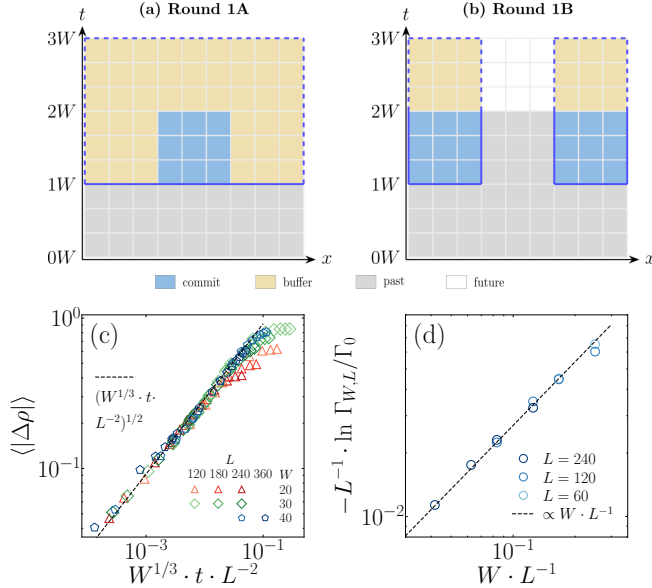


FIG. 3. (a, b) Layout of modular SWD. (c) Results for diffusive motion under modular SWD, compare with Fig. 2(a). (d) Results for the crossover scaling of the memory time under modular SWD, compare Fig. 1(c).

rounding as a buffer (Fig. 3(a)); and then commit to odd-numbered blocks, with their $W \times 2W$ surrounding as a buffer (Fig. 3(b)). Because the even blocks have already been committed, the odd blocks require a smaller buffer. Attention must be paid to the boundaries of the decoding windows: dashed lines between *buffer* and *future* indicates that the correction chain E' is allowed to cross that boundary, whereas solid lines between *commit* and *past* indicates that the correction chain E' is constrained not to cross that boundary. Similarities can be found between modular SWD and the modular decoder proposed in Ref. [23].

Our results are shown in Fig. 3(c,d). In Fig. 3(c), we find similar scaling of the diffusion constant D as in Eq. (3) and Fig. 2(a). Due to the lower computational cost, we can access larger system sizes while maintaining $W \ll L^\zeta$, so that the diffusive scaling can be verified for a larger range of W . In Fig. 3(d), we find similar crossover scaling of t_{mem} as in Eq. (5) and Fig. 1(c). By design, we are limited to regimes $W/L \ll 1$, as the blocks have spatial extent limited by the system size; this explains why we do not see the full range of the scaling function Φ .

We thus conclude that modular SWD correctly captures universal aspects of SWD, while requiring a smaller numerical overhead, as it operates on an $O(W) \times O(W)$ decoding window, rather than an $O(L) \times O(W)$ one. Due to its modularity, modular SWD should be the preferred choice over SWD for real-time decoding.

Union-Find and Clustering decoders

Next, we consider the 1D repetition code and SWD, but with MWPM replaced by either the union-find (UF) [55] or clustering [56, 57] decoders. Both algorithms proceed by growing clusters isotropically from seeds of point-like defects. Clusters stop growing whenever they become resolvable by an error chain, which in our convention are allowed to terminate on the boundary of a window. Comparing the two, the clustering decoder removes a resolvable cluster by finding and applying an error chain immediately after one is found, whereas the UF decoder performs global removal at the very end.

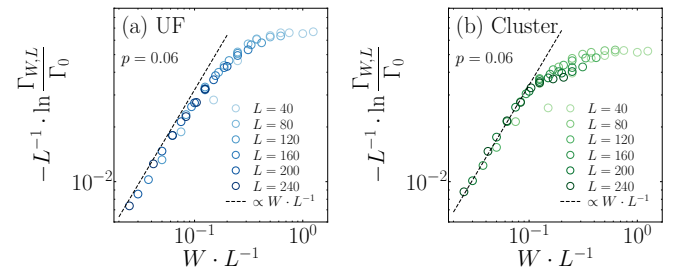


FIG. 4. Results from $\mathbb{P}_{\text{fail}}$ of SWD with the UF and clustering decoders. The data is in agreement with Eq. (5) and Fig. 1(c).

Under the UF or clustering decoders, pairs of $\mathbb{Z}_2$ charges separated by a distance greater than $O(W)$ are essentially invisible to each other. Therefore, similarly to SWD with MWPM, SWD with the UF or clustering decoders is also expected to undergo quasi-local motion as described by Eq. (1). Indeed, with numerical results in Fig. 4 we find agreements with the crossover scaling in Eq. (5) and Fig. 1(c). Note that the error rate is adjusted accordingly, as the UF and clustering decoders have a lower threshold.

Crossover scaling for the 2D toric code

We now examine whether the reaction-diffusion description established for the 1D repetition code in the main text extends to the 2D toric code. The code is defined on an $L \times L$ square lattice with periodic boundary conditions, and is subject to phenomenological noise: independent bit-flip (Pauli Z) errors on data qubits at rate p and independent measurement errors on syndrome bits at rate p . We decode using SWD with MWPM acting within each window, following the same commit/buffer/projection prescription as in Fig. 1(a).

The microscopic degrees of freedom differ from those of the 1D repetition code in a few ways. On each time slice, the syndrome is a collection of violated plaquette operators, i.e. point-like $\mathbb{Z}_2$ anyons. Therefore, $E \oplus E'$ is a collection of closed loops in (2+1)-dimensional space-time. (A similar picture holds for the phase-flip (Pauli X) errors.) The coarse-graining arguments nevertheless go through: syndrome pairs separated by more than W in spacetime are not resolved within a single decoding window, so the screening cutoff is still of order W , and the slow degrees of freedom of the effective theory are coarse-grained anyons undergoing parity-conserving reaction-diffusion process.

The main differences from the 1D case are that the anyons now diffuse in two spatial dimensions. Therefore, the scaling of the diffusion constant $D(W) \propto W^{z\zeta-1}$ will have a different value of ζ , given by a directed polymer

in three spatial dimensions. However, we have not been able to directly test this. The (1+1)D domain-wall order parameter $\Delta\rho$ used to track a coarse-grained charge in Fig. 2 has no direct (2+1)D analogue, and the position of an isolated coarse-grained charge cannot be read off from a local observable in the same way. For this reason, we do not present a test of the diffusion scaling for the toric code analogous to Fig. 2, and instead focus on the crossover scaling of the memory time.

Initializing the system in a logical codeword state we run SWD at fixed error rate $p < p_c$ for long enough to extract the logical failure rate $\Gamma_{W,L}$ from the linear growth of $\mathbb{P}_{\text{fail}}(t; W, L)$, as in the main text. We find that the same scaling form applies,

$$-L^{-1} \cdot \ln \frac{\Gamma_{W,L}}{\Gamma_0} = \Phi_{2D}(W/L), \quad (9)$$

where Φ_{2D} has the same asymptotic behavior as the 1D scaling function in Eq. (5), i.e. linear at small argument and saturating at large argument, see Fig. 5(b). We note that the static threshold for the 2D toric code under phenomenological noise is $p_c \approx 2.9\%$ [12, 58], lower than that of the 1D repetition code at the same abstract level of description; this limits the range of error rates over which the scaling can be tested deep in the decodable phase.

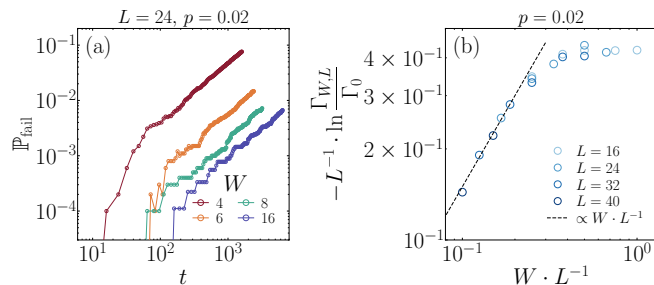


FIG. 5. Crossover scaling of the 2D toric code (compare with Fig. 1 for the 1D repetition code). For early times, we did not observe any logical failures in 10^4 realizations.

Supplemental Material for “Kinetics of sliding-window quantum error correction”

Adithya Sriram,¹ Charles Stahl,¹ Aleksander Kubica,² and Yaodong Li^{1,3}

¹*Department of Physics, Stanford University, Stanford, CA 94305*

²*Yale Quantum Institute & Department of Applied Physics, Yale University, New Haven, CT 06520*

³*Department of Mechanical Engineering & Department of Physics, University of Washington, Seattle, WA 98195*

(Dated: August 11, 2026)

S1. FORMAL DEFINITION AND EFFECTIVE DESCRIPTION OF THE SLIDING WINDOW DECODER (SWD)

In this section we describe the SWD in some more detail, and provide motivations for the effective description in Eq. (1).

A. Formal definition of the SWD

Our discussion is focused on d -dimensional topological CSS codes with deconfined point-like defects; see Ref. [12] for more details. In this setting, the decoding problem can be associated with a $(d+1)$ -dimensional spacetime lattice. The spacetime locations of physical errors and measurement errors are captured by the *error chain*; the *correction chain* captures the output of the local decoder $\mathcal{R}$, i.e., the locations of estimated errors. The detectors, which, in the absence of errors, are deterministic products of stabilizers, can be associated with the vertices of the spacetime lattice. The syndrome information corresponds to the triggered detectors (note that a detector is triggered only if the number of errors incident to it is odd). For the error chain E , we denote the corresponding syndrome as ∂E . For the specific cases of the repetition code and the d -dimensional toric code (with point-like defects), each individual error triggers at most two detectors, and therefore we can talk about $\mathbb{Z}_2$ charges at the endpoints of the error chain.

The task of the decoder is to output a correction consistent with the observed syndrome (or, in other words, a correction that annihilates the defects), while trying to minimize the probability of causing a logical error. This correction $E' = \mathcal{R}(\partial E)$ will naturally have common endpoints with E , namely $\partial E = \partial E'$; these endpoints ∂E correspond to “detection cells” (also known as frustrated plaquettes in an associated stat-mech model), which serve as inputs to $\mathcal{R}$. Their combination $E \oplus E'$ forms closed loops on the spacetime lattice, which for (1+1)D can be interpreted as domain walls of an effective Ising spin configuration. The closed loops in $E \oplus E'$ can also be viewed as worldlines of $\mathbb{Z}_2$ charges. Generally speaking, while the error chain E can be defined unambiguously and independently of any decoder, the correction chain E' must be defined with respect to a decoder.

The sliding window decoder (SWD) operates in rounds, labeled by a non-negative integer $n \in [0, N]$. Define the cumulative correction chain before round n to be

$$(E'_C)^{<n} := \bigoplus_{n' < n} E'_C(n'). \quad (\text{S1})$$

Here, $E'_C(n)$ is the *committed* correction chain in round n , to be defined inductively below. By convention, $(E'_C)^{(<0)} = \emptyset$ and $E'_C(n) = \emptyset$ for all $n < 0$.

During the n -th round, the decoder inputs syndromes within the thin strip in spacetime (see Fig. 1).

$$B(n) := [nW, (n+2)W] \times [0, L] \quad (\text{S2})$$

Within $B(n)$, the decoder performs decoding based on the observed detection cells as follows

$$E'(n) := \mathcal{R} \left\{ \partial \left[(E \oplus (E'_C)^{<n}) \cap B(n) \right] \right\}. \quad (\text{S3})$$

Notice that the syndrome is a result of both physical error E and past corrections $(E'_C)^{<n}$. Define the *commit window*

$$B_C(n) := [nW, (n+1)W] \times [0, L] \quad (\text{S4})$$

and the *buffer window*,

$$B_B(n) := [(n+1)W, (n+2)W] \times [0, L]. \quad (\text{S5})$$

The decoder only commits to corrections in $B_C(n)$,

$$E'_C(n) := E'(n) \cap B_C(n), \quad (S6)$$

whereas $B_B(n)$ serves as a look-ahead that improves decoding accuracy, but no final decisions are made there yet. As the commitment involves a change of basis on the physical space, the syndrome data will need to be transformed covariantly.

After committing to $E'_C(n)$, the decoder discards the syndrome data from $B_C(n)$, and $B_B(n)$ slides forward to become $B_C(n+1)$. Each round thus overlaps with its neighboring ones, and corrections are made every W time steps at a rate $R \propto 1/W$.

Special attention to boundary condition is necessary when performing $\mathcal{R}$ within $B(n)$. The correction chain $E'(n)$ might span both $B_C(n)$ and $B_B(n) = B_C(n+1)$, but the decoder will need to erase everything within $B_B(n)$ in the next round. Therefore, we must allow the possibility of *projecting* detection cells in $B_C(n)$ onto the temporal boundary between $B_C(n)$ and $B_B(n) = B_C(n+1)$, see Eq. (S6), where $E'_C(n)$ effectively implements this projection. In this way, the projected detection cells can be picked up in the next round, see Eq. (S3), where $\mathcal{R}$ receives input from $(E'_C)^{<n}$, including any such projection. This ensures that $E \oplus E'$ form a closed loop, from which a consistent error history can be reconstructed.

B. Quasi-local motion of coarse-grained $\mathbb{Z}_2$ charges

Viewing the closed loops in $E \oplus E'$ as worldlines of $\mathbb{Z}_2$ -charged point defects, we now argue that their motion admits a quasi-local description, as long as we look at length scales and time scales long compared to W .

A quasi-local description is not obvious *a priori*, as the decoder $\mathcal{R}$ performs a global matching of point-like detection cells, and $E'(n) := \mathcal{R}[\partial E \cap B(n)]$ can in principle be long-ranged. One might therefore expect that any effective description of the evolution of $E \oplus E'$ will necessarily be nonlocal as well.

However, because the detection cells are allowed to be paired with temporal boundary (see Fig. 1(a)), the window width W introduces a cutoff on the length of $E'(n)$. Indeed, consider a defect pair separated by $\Delta x \gg W$ in space. Within a single window, $\mathcal{R}$ can either match the two endpoints with each other (at cost $O(\Delta x)$) or match each endpoint locally with the temporal boundary (at cost $O(W)$). When $\Delta x \gg W$, the latter is always favorable, and the two endpoints are effectively “screened”. More generally, any connected component of length $\gg W$ in $E'(n)$ will be suppressed exponentially.

Therefore, we have argued that $E'(n) := \mathcal{R}[\partial E \cap B(n)]$ will typically have connected components of lengths at most $O(W)$. Therefore, $E'_C(n) \subseteq E'(n)$ also typically has connected components of lengths at most $O(W)$. Writing

$$\gamma_{\text{tot}} := E \oplus E' = \bigoplus_n \underbrace{[E \cap B_C(n)] \oplus E'_C(n)}_{=: \gamma(n)} \quad (S7)$$

and interpreting them as worldlines of $\mathbb{Z}_2$ charges, our argument implies that the $\mathbb{Z}_2$ charges also move in a quasi-local manner. In particular, the time evolution within the round n is determined by $\gamma(n)$, which contains local moves of size at most $O(W)$.

Nevertheless, the charges can still travel a long distance, if they survive over many rounds, see Fig. 1(b).

S2. BRANCHING-ANNIHILATING RANDOM WALKERS (BARW) UNDER SWAP NOISE CHANNEL

The reaction-diffusion description developed in the main text does not have an absorbing-state phase transition, because the noise in the standard SWD setting is always present and the empty configuration is never a strict fixed point of the dynamics. Here we describe a modified error model in which the absorbing state is indeed stable, allowing access to an absorbing state phase transition. This provides another check of our effective description.

We consider the 1D repetition code under a noise model in which bit-flip errors are turned off, and the only physical noise is a random SWAP channel acting on neighboring bits. A SWAP acts trivially on uniform configurations of all 0 or all 1, therefore preserve the absorbing states. In the language of Eq. (1), pair creation is forbidden ($\tau = 0$). When there is an isolated domain wall, a SWAP acting across the domain wall can move a 0 into the region of 1s (or vice versa), creating an isolated bit surrounded by its opposite and hence two new domain walls,

$$00001111 \rightarrow 00010111. \quad (S8)$$

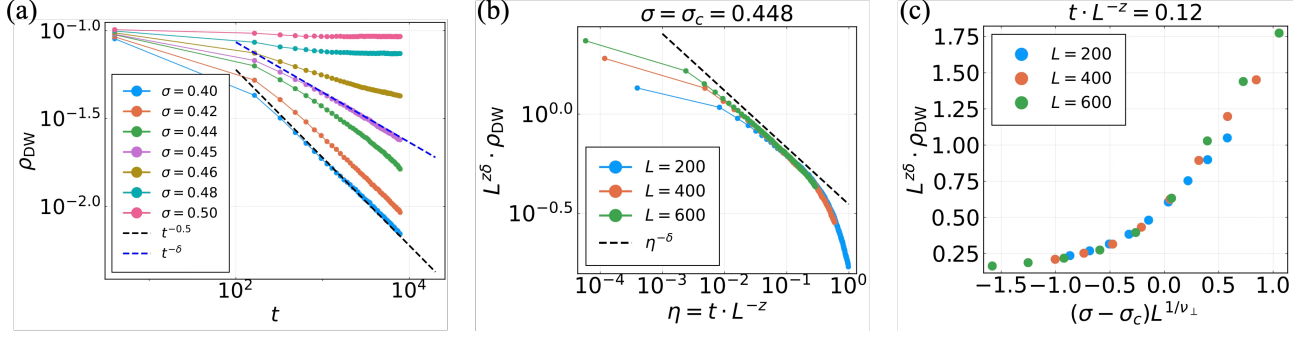


FIG. S1. Numerical results for the SWD under a SWAP noise channel, which we compare with a branching-annihilating process, see Eq. (S9). (a) When we increase the SWAP rate σ , the system goes from an *absorbing* phase where the domain wall density decays as $\rho_{\text{DW}} \propto t^{-1/2}$ at long times, to an *active* phase where ρ_{DW} saturates at a nonzero constant. (b, c) The critical data for ρ_{DW} can be fitted to Eq. (S10), and agreements with DP2 exponents are found.

Under the joint action of the SWAP channel (which supplies the 1-to-3 branching of domain walls) and the SWD feedback (which matches syndromes and contributes to annihilation on contact, as well as to diffusion), the domain walls evolve as a parity-conserving branching-annihilating random walk (BARW) [28, 59],



which is known to undergo an absorbing-state phase transition in the DP2 (parity-conserving) universality class as σ is tuned. Note that in the present setting D cannot be varied independently of σ .

We simulate this model and tune through an absorbing-state transition. We fix $W = 4$, take the SWD with MWPM as its inner decoder, and start with an initial state with domain wall density $\rho_{\text{DW}} = 0.2$. For $\sigma < \sigma_c$, the system is in the absorbing phase and we find the diffusion-limited asymptotic decay $\rho_{\text{DW}} \propto t^{-1/2}$, see Fig. S1(a). Near the critical point ρ_{DW} obeys the following standard scaling relation,

$$\rho_{\text{DW}}(\sigma; t, L) = t^{-\delta} \Upsilon(t \cdot L^{-z}, (\sigma - \sigma_c) \cdot L^{1/\nu_\perp}) = L^{-z\delta} \tilde{\Upsilon}(t \cdot L^{-z}, (\sigma - \sigma_c) \cdot L^{1/\nu_\perp}), \quad (\text{S10})$$

where $\delta \approx 0.285$, $\nu_\perp \approx 1.83$, $z \approx 1.74$ are exponents of the DP2 universality class [60]. The scaling ansatz as well as the DP2 exponents are consistent with our findings in Fig. S1(b,c). We view this as a further nontrivial test of our effective description, where the reaction-diffusion process is tuned across a critical point.

Our results should be generally applicable for any $W = O(1)$. The crossover to large W is an interesting question we leave for the future.