

Enhancing quantum memory lifetime with measurement-free local error correction and reinforcement learning

Mincheol Park ^{*}, Nishad Maskara , Marcin Kalinowski , and Mikhail D. Lukin[†]

Department of Physics, [Harvard University](https://www.harvard.edu), Cambridge, Massachusetts 02138, USA



(Received 30 August 2024; accepted 2 December 2024; published 13 January 2025)

Reliable quantum computation requires systematic identification and correction of errors that occur and accumulate in quantum hardware. To diagnose and correct such errors, standard quantum error-correcting protocols utilize *global* error information across the system obtained by mid-circuit readout of ancillary qubits. We investigate circuit-level error-correcting protocols that are measurement-free and based on *local* error information. Such a local error correction (LEC) circuit consists of faulty multi-qubit gates to perform both syndrome extraction and ancilla-controlled error removal. We develop and implement a reinforcement learning framework that takes a fixed set of faulty gates as inputs and outputs an optimized LEC circuit. To evaluate this approach, we quantitatively characterize an extension of logical qubit lifetime by a noisy LEC circuit. For the two-dimensional (2D) classical Ising model and four-dimensional toric code, our optimized LEC circuit performs better at extending a memory lifetime compared with a conventional LEC circuit based on Toom's rule in a subthreshold gate error regime. We further show that such circuits can be used to reduce the rate of mid-circuit readouts to preserve a 2D toric code memory. Finally, we discuss the application of the LEC protocol on dissipative preparation of quantum states with topological phases.

DOI: [10.1103/PhysRevA.111.012419](https://doi.org/10.1103/PhysRevA.111.012419)

I. INTRODUCTION

Quantum error correction (QEC) protects encoded quantum information from experimental noise and is critical to realizing the full computation power of quantum processors [1]. By embedding logical information into QEC codes, errors can be systematically identified and corrected [2]. In recent years, experimental progress has been rapid with realization of small codes and simple correction procedures being explored across various quantum hardware platforms, including the two-dimensional (2D) triangular color code [3] on trapped ions, the 2D surface code [4] on superconducting qubits, and the 2D toric code [5] and three-dimensional (3D) color codes [6] on neutral-atom arrays. However, the ultimate path to efficient and scalable error correction in large systems is far from clear. As such, a variety of approaches for reducing hardware requirements should be considered.

Stabilizer codes are a particularly important class, as they come equipped with efficient circuits for identifying and correcting errors without affecting the encoded information [7,8]. One approach to implement error correction with these codes is to use ancillary qubits and readouts to extract stabilizer eigenvalues. Then, stabilizer information is aggregated *globally* using a classical decoding algorithm, which identifies likely errors and implements corresponding recovery operations on the data qubits comprising the codes. These include decoders based on the minimum-weight perfect matching algorithm [9–16], machine learning [17–26],

or other approaches [27–30]. However, in principle, the correction procedure could be performed without intermediate communication with a classical processor, using coherent multi-qubit controlled gates and dissipation [31–37].

More specifically, there are measurement-free QEC protocols utilizing only *local* operations in its decoding procedure. These local error correction (LEC) strategies have received considerable theoretical attention, in particular in higher dimensions. Indeed, in dimensions $D \geq 4$, local cellular automata-based decoders have reported thresholds under phenomenological noise models, indicating the possibility of self-correcting quantum memories [38–43]. Moreover, efforts have extended into optimizing local decoding protocols by training convolutional neural networks [44,45], and practical implementation and application of local decoders in near-term noisy quantum systems [46–51]. Nevertheless, designing LEC circuits with a restricted faulty multi-qubit gate set is a challenging task. For this purpose, it is desirable to develop a systematic optimization framework that enables such circuit design.

Reinforcement learning (RL) is a powerful framework for optimizing sequences of operations to perform a particular task [52]. As such, RL has been applied recently in quantum science and engineering [53] to develop both quantum error-correcting codes [54–56] as well as decoders [57–60]. RL was also utilized to optimize noisy quantum circuits or controls that perform variational quantum algorithms [61–63], logical quantum gates [64], and state preparation [65–68]. For the state-of-the-art or near-term quantum hardware with few physical qubits, RL was applied to optimize quantum circuits that preserve quantum memory [69] and prepare logical states [70].

^{*}Contact author: mincheol_park@g.harvard.edu

[†]Contact author: lukin@physics.harvard.edu

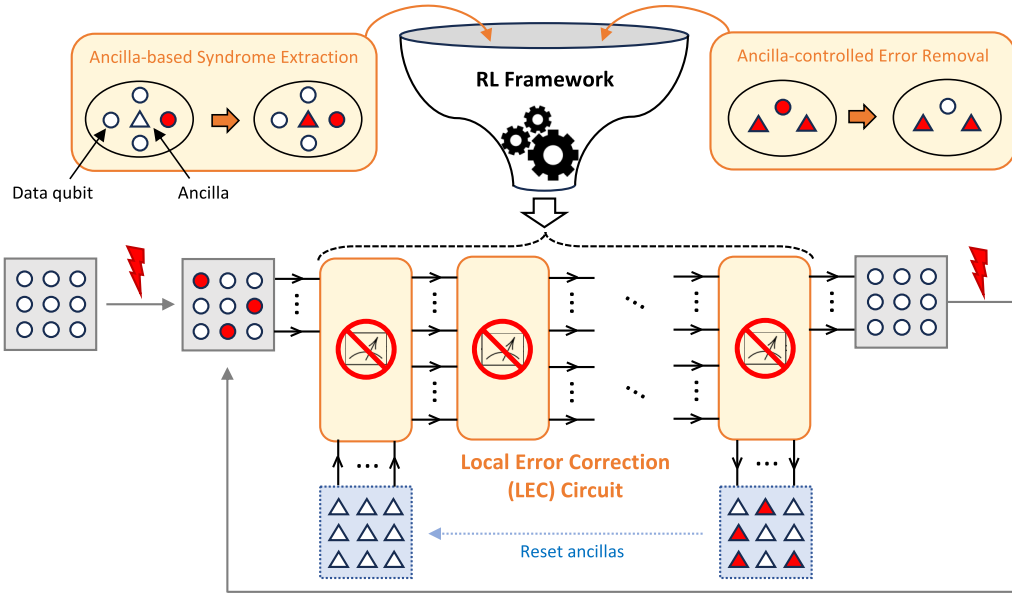


FIG. 1. Overview of circuit-level local error correction (LEC) protocol and the reinforcement learning (RL) optimization framework. The LEC circuit is a sequence of local coherent quantum operations—not including a mid-circuit projective readout of any qubit—applied to data qubits (circle) and ancillas (triangle). It takes data qubits with errors (colored red) and initialized ancillas as input and moves errors of data qubits into ancillas. This local decoding can be repeated over multiple cycles with a finite number of ancillas by resetting them into the initialized ones. A reinforcement learning (RL) framework constructs an optimized LEC circuit by combining two types of “building block” operations, followed by corresponding coherent and independent reconfiguration of qubits. First, *ancilla-based syndrome extraction* operations enable the ancillas to measure and store syndrome information that indicates the existence of errors on neighboring data qubits. On the other hand, *ancilla-controlled error removal* operations provide local feedback operation controlled by ancillas onto data qubits to correct their errors. All these operations can be achieved experimentally within the state-of-the-art quantum processors, including the Rydberg atom arrays (see Appendix A).

In this work, we optimize circuit-level LEC protocols that consist of faulty multi-qubit gates—to perform both syndrome extraction and ancilla-controlled error removal—by developing an RL framework. In particular, our RL framework takes a fixed set of faulty gates as inputs and outputs an LEC circuit, with optimized length and layout. By performing extensive circuit-level simulations, we show these circuits successfully reduce the error density when gate fidelities are high enough. We then characterize the resulting extension in the lifetime of a memory encoded in three finite-sized systems: 2D toric code, 2D Ising model, and four-dimensional (4D) toric code. Finally, we explore two near-term applications of this scheme: (1) reducing the rate of mid-circuit readouts required to preserve quantum memory and (2) dissipatively preparing quantum states of topological phases.

This paper is organized as follows: we begin in Sec. II by outlining the key insights and main results of this work. Section III provides a detailed analysis of a circuit-level LEC scheme. In particular, we present an implementation of the basic components of the LEC circuits followed by characterization and resolution of the error patterns that limit conventional LEC circuits. Then, Sec. IV discusses a setting for the RL framework and its optimization of circuit depth and layout within each code. Section V presents a quantitative characterization of memory lifetime extension by LEC circuits and a comparison between RL-optimized and conventional LEC circuits. Furthermore, Sec. VI proposes two near-term applications of this scheme: reducing the rate of mid-circuit readouts required to preserve quantum memory

in 2D toric code and dissipatively preparing quantum states of topological phases. Finally, we present conclusions and outlook in Sec. VII.

II. OVERVIEW OF MAIN RESULTS

We focus on a circuit-level, measurement-free LEC protocol, composed of faulty multi-qubit gates. As illustrated in Fig. 1, the circuits are constructed from two types of basic local quantum operations: (1) ancilla-based syndrome extraction with a series of two-qubit gates and (2) ancilla-controlled error removal with three-qubit gates. We focus on three finite-sized error-correcting codes—the 2D toric code, classical 2D Ising model, and 4D toric code—and start by studying the performance of existing LEC approaches, such as nearest-neighbor matching in 2D toric code [46] or Toom’s rule in the 2D Ising model and 4D toric code [38]. We then develop and apply an RL framework to optimize LEC circuits and improve the overall performance in all three cases. Finally, we provide a detailed and intuitive understanding of how the optimization achieves improved performance.

We find that the existing LEC methods are limited by specific low-weight error patterns, which are not removed by the local error-removal operations comprising the circuit, as also noted by previous literature [40,44]. To address this, we introduce additional longer-range operations that are capable of removing some of these error patterns and thus increase the weight of the minimum uncorrectable error in each code. The RL framework, as shown in Fig. 6, assembles

the operations into optimized LEC circuits that prevent the accumulation of higher-weight error patterns on average over multiple rounds of their application. In general, since we are considering a circuit-level noise model, adding additional gates also increases the number of errors. Importantly, we find that the RL can balance this cost, with the benefits arising from removing such errors: as gates get noisier, the optimization prefers shorter and simpler LEC circuits to minimize errors. In contrast, when we consider gates with higher fidelity, the RL procedure uses more complex decoding procedures to further improve performance.

We illustrate the utility of RL-optimized LEC circuits by showing how they can extend the lifetime of quantum memory in a subthreshold-gate error regime. In each LEC round, we introduce “ambient” errors, which model noises arising from other faulty operations including idling or transversal logical gates [6,71], with p_{amb} error rate. Then, we apply a full LEC circuit that consists of faulty multi-qubit controlled gates with p_{gate} error rate. We define an average logical qubit lifetime T to be an average number of LEC rounds before encountering the first failure in decoding. Simulating different values of p_{amb} , we observe a scaling of the error suppression consistent with a power-law dependence upon T and p_{amb} (see Fig. 9). To quantify this effect, we introduce the notion of an *effective code distance* D_{eff} as a function of a given LEC circuit and p_{gate} such that $T \sim (p_{\text{amb}})^{-D_{\text{eff}}}$. D_{eff} is related to the ability of a noisy decoding circuit to remove (on average) error patterns of a certain weight. In the subthreshold p_{gate} regime, we numerically show that RL-optimized LEC circuits have higher D_{eff} compared with conventional LEC circuits in each code and are thus more effective in removing errors from noisy states (see Fig. 10).

Finally, we explore two potential applications of our LEC method. First, we explore how RL-optimized LEC circuits can be interleaved with standard (global) decoding, to reduce the number of mid-circuit measurements required to achieve a target performance, which could be used to e.g., reduce cycle times in logical processors. For example, we show that using LEC circuits with high-fidelity multi-qubit gates can reduce the number of required mid-circuit readouts (see Fig. 11). Second, noting that ideas from LEC have also been used to rigorously verify nontrivial, topological phases of matter in the presence of experimental noises [72,73], we show that our LEC scheme can similarly be applied in the dissipative preparation of topological phases with a given Hamiltonian.

III. CIRCUIT-LEVEL LOCAL ERROR CORRECTION SCHEME

In this section, we overview our family of circuit-level LEC schemes in 2D toric code, classical 2D Ising model, and 4D toric code by providing an explicit translation into a quantum circuit with multi-qubit gates from the local decoding rule [38,46,74,75]. Section III A discusses how to implement the basic components of the LEC circuit. Section III B characterizes the error patterns that limit conventional LEC circuits and then introduces the hierarchically extended set of error-removal operations to address this limitation of the circuit-level LEC scheme. To provide relevant details, we include the considerations of the experimental implementation

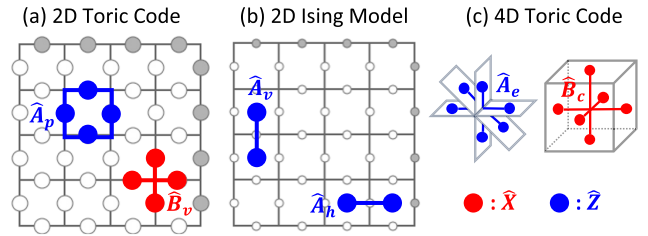


FIG. 2. Data qubits and stabilizers for each error-correcting code. Pauli \hat{X} and \hat{Z} operators on qubits at corresponding lattice positions are colored red and blue, respectively. Also, gray qubits denote a periodic boundary condition of a lattice. (a) In the 2D toric code lattice, data qubits live at edges. Stabilizer \hat{A}_p and \hat{B}_v are defined at each plaquette p and vertex v , respectively, by the tensor product of four \hat{Z} s and four \hat{X} s, respectively, on adjacent edges. (b) In the 2D Ising model lattice, data qubits live at plaquettes. Stabilizer \hat{A}_v and \hat{A}_h are defined at each edge by the tensor products of two \hat{Z} s on adjacent plaquettes. (c) In 4D toric code lattice, data qubits live at faces. Stabilizer \hat{A}_e and \hat{B}_c are defined at each edge e and cube c , respectively, by the tensor products of six \hat{Z} s and \hat{X} s, respectively, on adjacent faces.

of the LEC scheme in Appendix A and implementation of a code base outlined in this section in Appendix B.

LEC circuits are constructed as sequences of *local* coherent quantum operations between data qubits and ancillas (see Fig. 1). Importantly, there is no need for readout and classical processing in the circuit. Conceptually, the circuit takes noisy data qubits and noise-free ancillas as input and reduces errors on the data qubits by moving them to the ancillas through entangling operations. Thus, the LEC circuit effectively moves entropy from data qubits onto the ancillas. In our formulation, we distinguish errors arising from the LEC correction and errors arising from other sources, such as idling errors or other logical-circuit elements. As such, after each application of the LEC circuit, a new set of independent and identically distributed (i.i.d.) errors are applied to data qubits. The ancillas are then reset or discarded and the process is repeated, forming a *cycle* of LEC.

As depicted in Fig. 1, the RL framework functions as a “funnel” that takes the set of available local operations as input and produces a sequence composed of these elements, with optimized length and layout. We discuss the RL framework in more detail in Sec. IV. The input operations naturally decompose into two functions:

- (1) *ancilla-based syndrome extraction* by encoding stabilizer operators onto ancillas without readout,
- (2) *ancilla-controlled error removal* by applying local feedback controlled by ancillas onto data qubits.

Each of these operations requires applying multi-qubit gates to the system *in parallel* following a desired layout.

A. Circuit-level local error correction operations in each code

We use three models to develop and test our framework, illustrated in Fig. 2. The 2D toric code is a canonical quantum error-correcting code, which has been studied extensively [9,10]. This code, having point-like syndrome excitations, cannot scalably protect quantum information with

a local decoding protocol. Nevertheless, we still observe pseudothreshold behavior, as discussed further below, and gain simple intuition for how LEC works using this model. To extend our analysis to models that can, in principle, be protected by an LEC procedure, we consider the 2D (classical) Ising model, which realizes a classical memory [76], and the 4D toric code, which realizes a quantum memory [9,77]. Both have line-like syndrome excitations, which can be corrected via local operations [38,74,75,78]. With sufficiently low error rates, LEC circuits can protect their information indefinitely, making these codes *self-correcting memories*. Each model can be defined by its stabilizers, where their eigenvalues are either $+1$ or -1 . Then, each model encodes logical qubit(s) where all stabilizers have $+1$ eigenvalues. Note that each lattice of the error-correcting code below has a periodic boundary condition.

(1) 2D toric code [79]

As in Fig. 2(a), data qubits live on edges of the 2D lattice, and plaquette-type (\hat{A}_p) and vertex-type (\hat{B}_v) stabilizers are tensor products of four Pauli \hat{Z} s (or \hat{X} s) on data qubits neighboring each plaquette p (or vertex v). Two logical qubits are encoded with logical \hat{Z} (or \hat{X}) that correspond to tensor products of \hat{Z} s (or \hat{X} s) along the topologically nontrivial loops.

(2) 2D Ising model [9]

As in Fig. 2(b), data qubits live on plaquettes of the 2D lattice, and vertical-type (\hat{A}_v) and horizontal-type (\hat{A}_h) stabilizers are tensor products of two \hat{Z} s on the data qubits vertically or horizontally adjacent to each edge. Since the 2D Ising model can detect only a bit-flip error on the qubit, it encodes a classical memory. Also, one logical qubit is encoded that consists of all qubits with either $|0\rangle$ or $|1\rangle$.

(3) 4D toric code [9]

As in Fig. 2(c), data qubits live on the faces of the 4D lattice, and there are edge-type (\hat{A}_e) and cube-type (\hat{B}_c) stabilizers. Each stabilizer is a tensor product of six \hat{Z} s (or \hat{X} s) on data qubits neighboring each edge e (or cube c). Two logical qubits are encoded with logical \hat{Z} (or \hat{X}) that correspond to tensor products of \hat{Z} s (or \hat{X} s) on the topologically nontrivial sheets.

The syndrome extraction circuits are used to measure stabilizers via ancillas. For each of the above error-correcting codes, the stabilizers are a tensor product of \hat{Z} s or \hat{X} s on the data qubits. Since \hat{Z} and \hat{X} anticommute with each other, eigenvalue of stabilizer \hat{S}_Z (or \hat{S}_X) is $(-1)^{n_s}$, where n_s is a number of \hat{X} (or \hat{Z}) errors on data qubits that support the stabilizer. Due to the commutation relation between a controlled-NOT (CNOT) gate and \hat{X} and \hat{Z} , a sequence of CNOT gates performs a syndrome extraction. For example, for 2D toric code, we put an ancilla initialized to $|0\rangle$ on every plaquette and then apply the sequence of parallelized CNOT gates as shown in Fig. 3(a). Then, each ancilla remains as $|0\rangle$ if an eigenvalue of corresponding \hat{S}_Z is $+1$ and flips into $|1\rangle$ otherwise. Similar sequences of CNOT gates in parallel can detect errors by syndrome extraction for the 2D Ising model and 4D toric code. Note that we assume a reset of ancillas to their initial states $|0\rangle$ before a new syndrome extraction operation.

The error-removal operations are more complex and involve applying correction gates conditioned on the state of the ancillas. Let $|A_1\rangle$ and $|A_2\rangle$ be two ancillas that store an

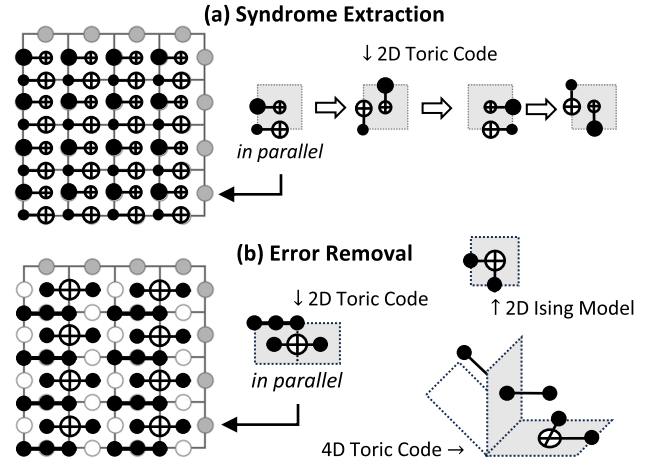


FIG. 3. The gate-level implementation of syndrome extraction and error-removal operation in each code. (a) Syndrome extraction operation using ancillas is a sequence of CNOT gates in parallel. An example of such a sequence with unit cells for 2D toric code is shown. Note that the CNOT gates are applied in the optimized order (see Appendix B). (b) Error-removal operation to reduce error density of data qubits controlled by ancillas is a single application of CCX and CCZ gates in parallel. An example of the simplest such operation with a unit cell for each code is shown. To remove \hat{X} (or \hat{Z}) errors, CCX (or CCZ) gates are applied.

eigenvalue of stabilizer \hat{S}_{Z,A_1} and \hat{S}_{Z,A_2} , respectively. Then, applying a controlled-controlled- X (CCX) gate controlled by these two ancillas—without knowing their quantum states—and targeted on the data qubit $|D\rangle$ results in $\hat{X}|D\rangle$ if both of the controlled qubits are $|1\rangle$ s and $|D\rangle$ otherwise. As the simplest scenario, consider these two stabilizers share one supporting data qubit $|D\rangle$. If both stabilizers have -1 eigenvalues, we have $|A_1\rangle = |A_2\rangle = |1\rangle$ by syndrome extraction and there are an odd number of errors on data qubits supporting each of these stabilizers. When data qubits have low error density, it is likely that the shared data qubit $|D\rangle$ has an \hat{X} error, which gets corrected by applying the above CCX gate. We can similarly remove a \hat{Z} error on the data qubit by applying a controlled-controlled- Z (CCZ) gate. The simplest gate layout for an error-removal operation in each code is shown in Fig. 3(b).

A typical *global decoding* approach also uses ancilla-based syndrome extraction operations. However, instead of performing the error-removal operation via local three-qubit gates, all the ancillas are projectively measured after the syndrome extraction operation. Then, by globally processing readout outcomes, a classical decoding algorithm infers the most likely underlying errors, which can be subsequently corrected by applying local Pauli operators. However, in our LEC scheme, only certain local correction operations are implemented, leading to the existence of additional *uncorrectable* error configurations.

B. Extended local error correction operations

The syndrome extraction and error-removal circuits form the building blocks of LEC circuits. To understand the capabilities of LEC, we start by discussing how different kinds of

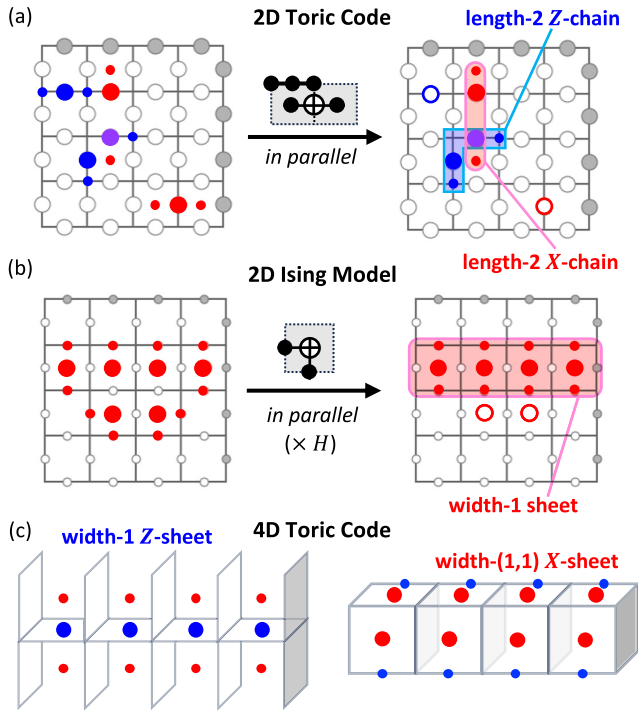


FIG. 4. Uncorrectable error pattern by the simplest error-removal operation in each code, as introduced in Fig. 3(b). (a) In 2D toric code, the given error-removal operation removes only length-1 error chains that consist of a single \hat{X} or \hat{Z} error on a data qubit. However, error chains with more than one error are uncorrectable by the given operation. (b), (c) In the 2D Ising model and 4D toric code, illustrated membrane-like error sheets bounded by parallel 1D syndrome lines that are periodic around the lattice are uncorrectable by the arbitrary repetition of Toom's rule error-removal operation.

errors behave under these circuits. Each code has a specific error pattern where syndromes, which are -1 eigenvalues of stabilizers, form a boundary of errors on neighboring data qubits. As shown in the left of Fig. 4(a), 2D toric code has 1D chains of errors on data qubits, which share a common adjacent plaquette or vertex, with endpoints of syndromes. These error chains can be classified by their lengths, i.e., how many errors on data qubits are between two syndrome endpoints to compose the chains. Here, we define the length of an error chain based on the shortest chain up to multiplication with stabilizers. Applying the simplest error-removal operation can remove length-1 X -chains or Z -chains that consist of a single \hat{X} or \hat{Z} error on a data qubit [46].

In contrast, as shown in the left of Figs. 4(b) and 4(c), the 2D Ising model and 4D toric code have 2D membranes of data qubit errors with 1D boundaries of syndromes. For these codes, repeating Toom's rule operation is an established cellular automata decoding protocol [74,75]. This operation can be implemented with CCX and CCZ gate as depicted in Fig. 4(b) followed by syndrome extraction. As this operation effectively "shrinks" the size of a 2D error membrane, its repetition can remove all errors in an infinite-sized system if the error density of data qubits is below some threshold [72,78,80]. Due to this property, the 2D Ising model and 4D toric code are called *self-correcting memories*.

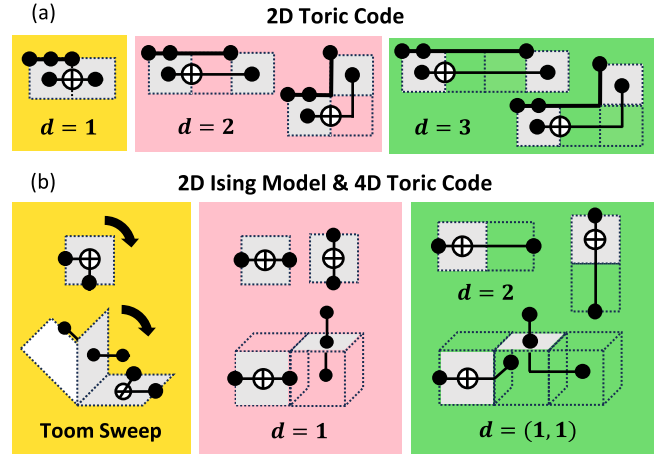


FIG. 5. Unit cells of extended error-removal operations for each code. To address the uncorrectable error patterns shown in Fig. 4, we hierarchically introduce new error-removal operations. (a) In the 2D toric code, we include $d = N$ operations, for $N \leq 3$, that can reduce length- N chains into length- $(N - 1)$ chains. (b) For the 2D Ising model and 4D toric code, we extend Toom's rule operation to *Toom sweep* operations as generated by 90° rotations of the original Toom's rule operation. Additionally, we introduce $d = N$ operations, for $N \leq 2$ in the 2D Ising model and for $N = 1$ and $(1,1)$ in 4D toric code, which can reduce width- N sheets into width- $(N - 1)$ sheets. Note that $d = (1, 1)$ operation for 4D toric code can reduce width- $(1,1)$ sheets into width-1 sheets.

However, we observe error patterns that are *uncorrectable* by the simplest error-removal operation, as shown in Fig. 4. In 2D toric code, the simplest error-removal operation cannot remove nor reduce the error chains with length ≥ 2 . Also, for the finite-sized 2D Ising model and 4D toric code, some error configurations are uncorrectable by repeated Toom's rule operations—even if the error density of data qubits is below the threshold. These configurations, also called as *energy barriers* [44,81,82], form periodic 2D sheets of data qubit errors surrounded by *parallel* one-dimensional (1D) loops of syndromes. We call these error patterns *error sheets* with width defined as the distance between the boundary 1D syndrome loops. Therefore, each error-correcting code has an issue of error patterns uncorrectable by the simplest error-removal operation arising from different reasons: the operation's limitation for 2D toric code and a system's finiteness for the 2D Ising model and 4D toric code.

Now, we introduce how to resolve this issue of uncorrectable error patterns that limit the performance of the local decoding protocol. We design a set of operations for each code such that any single low-weight error pattern can, in principle, be corrected.

For 2D toric code, we categorize the simplest removal operations as $d = 1$ operations, because they can reduce length-1 error chains into length-0 error chains, i.e., no error. As shown in Fig. 5(a), we generalize this to $d = N$ operations that can reduce length- N error chains into length- $(N - 1)$ error chains. We include up to $d \leq 3$ operations in our optimization.

In the 2D Ising model and 4D toric code, we extend the simplest operations in two ways, as shown in Fig. 5(b).

First, we include the Toom sweep operations generated by 90° rotations of the Toom operations, which were shown to reduce the occurrence of uncorrectable membrane-like error patterns [40,41]. Moreover, to enable the removal of such error patterns, we additionally introduce operations designed to remove errors between the two parallel 1D boundaries of syndromes. These operations take the same form as the $d = N$ operations for 2D toric code and can reduce width- N error sheets to have narrower widths. We include two lowest-weight operations for both codes— $d = 1$ and $d = 2$ operations for the 2D Ising model and $d = 1$ and $d = (1, 1)$ operations for the 4D toric code.

If we did read out the syndromes, then we would apply the proper combination of these operations within the extended set to correct any error chain with length ≤ 3 or any error sheet with width ≤ 2 and $\leq (1, 1)$. However, because the correction steps are applied via noisy controlled gates without the explicit knowledge of syndromes, constructing high-performance sequences of operations is quite challenging. To overcome this challenge, we develop an RL framework to optimize LEC circuits, which are sequences of the error-removal operations as well as syndrome extraction operations.

IV. REINFORCEMENT LEARNING OPTIMIZATION FRAMEWORK

In this section, we provide details about the RL framework to optimize LEC circuits. Section IV A introduces the setting of the RL framework in terms of its agent, environment, and the interplay between them. Then, Sec. IV B discusses the result of depth and layout optimization of LEC circuits by the RL framework with varying multi-qubit gate fidelities.

RL is a machine-learning paradigm where an *agent* learns which *actions* to play to maximize the *reward* through its interplay with an *environment* [52,83]. In our setting, the LEC circuit builder (or agent) learns which sequence of LEC gate operations (or actions) maximizes the decoding success rate (or reward) through its interplay with the circuit-level simulator (or environment) until the convergence of the agent-built LEC circuit. Due to the absence of mid-circuit readout in our scheme, our environment does not provide the *observation* to the agent, unlike the traditional RL setting [52,83].

We aim to optimize an ordered sequence of discrete actions where an optimizer only observes the reward of the full sequence. Since the sequence is not adjusted during the execution and the reward structure is “opaque” to the optimizer, our setting can be classified as an *open-loop control* or *black-box sequential optimization*. Our problem features (1) a large search space of size between 10^{55} and 10^{85} possible ordered sequences, (2) the importance of discrete actions’ order in the sequence, and (3) costly reward computation—especially for 4D toric code.

Due to these features, we choose a proximal policy optimization (PPO) among RL techniques as our optimizer [84]. The RL algorithms have been used to robustly and sample-efficiently optimize the sequence of actions in large discrete action spaces [85–88]. In particular, such algorithms have been successfully applied to the open-loop control problems [89–91]. More recently, the PPO algorithm has been used to optimize the quantum gate sequence, such as preparing

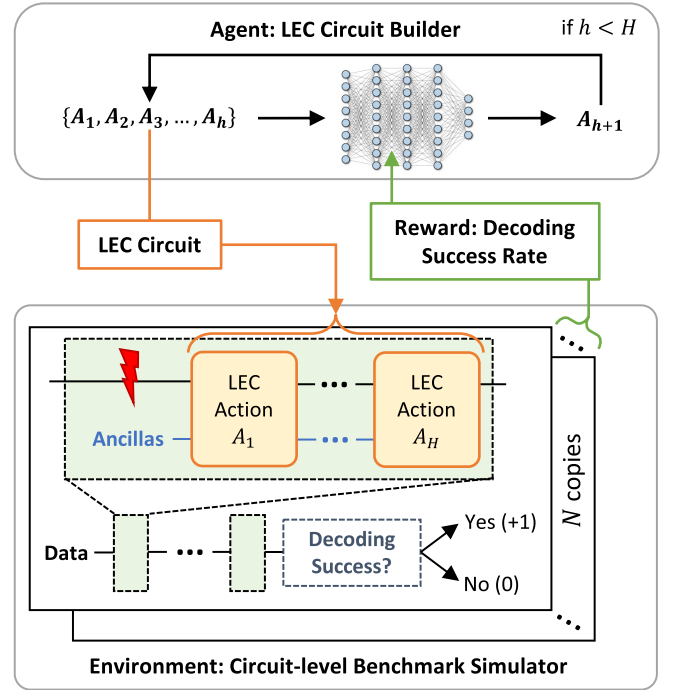


FIG. 6. Overview of our reinforcement learning (RL) framework that consists of an agent, environment, and their interactions to optimize the LEC circuit for a given code and error model. By choosing the next action based on all previous actions, the agent’s policy network builds an LEC circuit, which is a sequence of fixed circuit depth H number of LEC actions, as the environment’s input. Then, the environment benchmarks a decoding success rate given this LEC circuit with a fixed error model by simultaneously evaluating the success or failure of decoding for N copies of codes after multiple LEC cycles, which are applications of LEC circuits after the introduction of new errors. This rate is then used as a reward to update the agent’s neural network parameters. These steps are repeated until the perceived reward saturates, i.e., until the agent receives the same reward signal and does not design a new LEC circuit for multiple network updates.

multi-qubit GHZ states [92] or fault-tolerant logical states in small quantum error-correcting codes [70]. Although there are alternative optimizing strategies including variations of genetic algorithms [93] or simulated annealing [94], we find that the RL method is particularly suitable considering the purpose and features of our problem.

A. Setting of reinforcement learning framework

We summarize the specific interplay between the RL agent and environment as in Fig. 6 and Algorithm 1.

First, the RL agent produces an LEC circuit based on its internal parameters. Then, the RL environment takes this LEC circuit as an input to compute its decoding success rate with fixed code and error model parameters. Such a rate is used as a reward to update the internal parameters of the RL agent. This training process is repeated to maximize this reward until the resulting LEC circuit converges.

Let us further discuss each component of our RL framework: an agent, environment, and algorithm updating the

ALGORITHM 1. RL training algorithm.

Data: Circuit depth H ; action space \mathcal{A} ; code type and parameters; error model parameters
Result: RL-optimized LEC circuit
 LEC circuit \leftarrow EmptyList;
 Initialize parameter $\vec{\theta}$ of neural network $f^{\vec{\theta}}$;
repeat
 LEC circuit = BuildLEC($f^{\vec{\theta}}$, H , \mathcal{A});
 Reward = ComputeReward(LEC circuit, code type and parameters, error model parameters);
 $\vec{\theta} :=$ UpdateNetworkParameters($f^{\vec{\theta}}$, Reward);
until LEC circuit stays the same for multiple updates;

agent's neural network based on reward from the environment:

(1) The agent has a neural network $f^{\vec{\theta}}$ with vectorized internal parameters $\vec{\theta}$ that takes a sequence of h (for $h < H$) actions as input and returns the next action as output, i.e.,

$$f^{\vec{\theta}} : \mathcal{A}^{\otimes h} \rightarrow \mathcal{A}, \quad (1)$$

where \mathcal{A} denotes a set of all action candidates that depend on a choice of the error-correcting code. Repeating such next-action decision-making for H times results in building the LEC circuit with the circuit depth H . See Appendix C 1 for more details.

(2) The environment computes the decoding success rate by a circuit-level Monte Carlo simulation of performing multiple rounds of applying an LEC circuit produced by the RL agent. This simulation takes a code type, error model parameters, and LEC circuit as inputs. Here, we consider two different error sources: (1) *ambient* errors introduced every LEC round before applying the LEC circuit with the rate p_{amb} and (2) *gate* errors introduced after applying each parallelized multi-qubit gate with rate p_{gate} . Note that the reward computation is based on the *average* performance of an LEC circuit produced by the RL agent, because the decoder cannot access syndrome information in our measurement-free scheme. See Appendix C 2 for more details.

(3) We utilize the Python-based package Stable-Baselines3 to perform the training to implement the RL framework [95]. Recall that we update the agent's network parameters from the reward with the PPO algorithm. To obtain the best-performing LEC circuit, we run the RL training four times independently until the training reaches its termination condition. Among the final LEC circuits from four trained models, we choose the one that maximizes the reward.

Appendix C includes more details on our RL framework: a training termination condition (see Appendix C 3), hyperparameters (see Appendix C 4), and stability of training (see Appendix C 5).

B. Optimizing depth and layout of local error correction circuit

In each error-correcting code, we provide varying gate error rates p_{gate} as an input to our RL framework to optimize the length and layout of LEC circuits for each input p_{gate} . We choose the scale of p_{gate} to be with an order of 10^{-4} for 2D

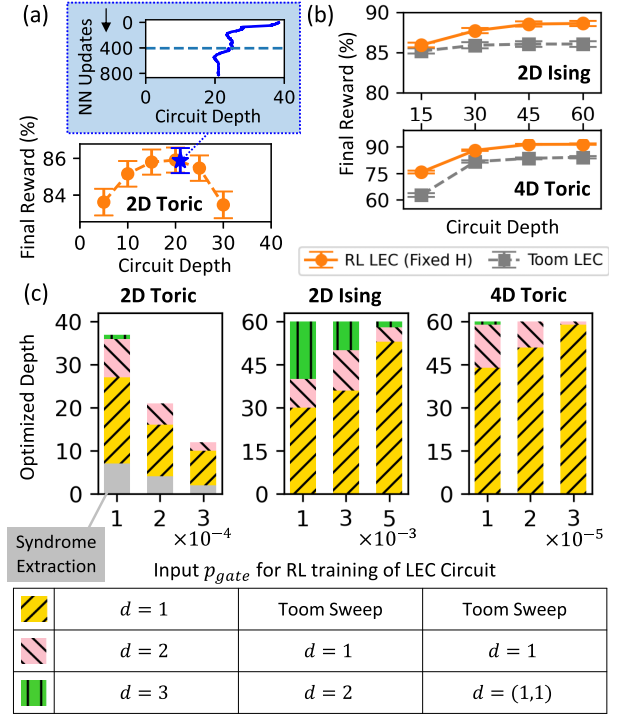


FIG. 7. Optimization of depth and layout of the LEC circuit by the RL framework in each error-correcting code. Each final reward is computed with 10 000 independent samples (see Appendix D for the error bar). (a) For 2D toric code, we show the RL training with variable depth with training $p_{\text{gate}} = 2 \times 10^{-4}$: as number of neural network (NN) policy updates increases, the LEC circuit's variable depth converges. By comparing the final reward of the resulting LEC circuit trained by the RL framework with fixed vs variable depth setting, we confirm that the obtained circuit depth is effectively optimized. (b) For the 2D Ising model and 4D toric code, we show the final reward after RL training vs fixed circuit depth with training $p_{\text{gate}} = 3 \times 10^{-3}$ for the 2D Ising model and 2×10^{-5} for 4D toric code. (c) For each training p_{gate} , we classify the components of RL-optimized LEC circuits as categorized in Fig. 5.

toric code, 10^{-3} for the 2D Ising model, and 10^{-5} for 4D toric code (see Table IV in Appendix C 2 for other code and error model parameters).

From RL training with varying circuit depth H , we observe that there exists an *optimal* circuit depth for 2D toric code, as shown in Fig. 7(a). As shown in Fig. 7(c), we observe that the optimized circuit depth H_{opt} for 2D toric code depends on training p_{gate} : H_{opt} decreases as training p_{gate} increases. This trend is because faulty LEC action causes a trade-off between error removal by parallel three-qubit gate operation and error introduced by a gate error model. Even in $p_{\text{gate}} = 0$ case, error removal by h th LEC action A_h gets marginalized as h increases. When p_{gate} increases, A_h for large h introduces more errors than it removes. We further observe that higher-weight LEC actions ($d > 1$) are more frequently used for smaller training p_{gate} . This trend is because higher-weight error chains are less common than lower-weight error chains in the lattice and thus lower-weight LEC actions benefit more from the trade-off of faulty LEC actions for smaller training p_{gate} .

Unlike in 2D toric code, $H > H_c \approx 45$ shows a saturation of final reward in the self-correcting memories as shown in Fig. 7(b). This is consistent with the system entering a steady state, where an action coming after some H_c achieves a balance between errors that it newly introduces by finite gate error or removes. Note that as $H \rightarrow \infty$, we expect the reward to slowly go to zero due to the rare possibility of high-weight uncorrectable errors. Thus, for the 2D Ising model and 4D toric code, we use the RL framework that optimizes circuit layout with fixed $H = 60$ (see Appendix C 1 for details). However, just as in 2D toric code, Fig. 7(c) also shows that new LEC actions ($d = N$) are more frequently used for smaller training p_{gate} with the same reasoning. Moreover, in both the 2D Ising model and 4D toric code, these new LEC actions appear only after a repetition of multiple Toom sweep actions. For all training p_{gate} , RL-optimized LEC circuits start with ≥ 10 Toom sweep actions for 2D Ising model and ≥ 19 Toom sweep actions for 4D toric code. This is because the membrane-like uncorrectable error configurations illustrated in Fig. 4 appear only after applying an iteration of Toom's rule decoding.

V. MEMORY LIFETIME EXTENSION BY LOCAL ERROR CORRECTION

In this section, we provide a thorough analysis of the memory lifetime extension by the LEC circuits and the comparison between RL-optimized LEC circuits (or *RL LEC*) vs conventional LEC circuits (or *conventional LEC*) in each code. Recall that the conventional LEC for 2D toric code is called a nearest-neighbor LEC circuit (or *NN LEC*) [46] and that for 2D Ising model and 4D toric code is called a Toom's rule-based LEC circuit (or *Toom LEC*) [38]. Section V A explains the idea of effective code distance that characterizes the lifetime of a memory encoded in the error-correcting code. Then, Sec. V B provides a comparison of this parameter for RL vs conventional LEC based on the benchmark simulation results. To explain why RL LEC performs better than conventional LEC, we provide the error statistics analysis in Appendix E 1.

A. Idea of effective code distance

We define the *average memory lifetime* T encoded in each code to be an average number of LEC rounds before the final recovery step calls decoding failure for the first time. We now introduce a parameter that represents the faulty LEC circuit's performance in extending the memory lifetime—called *effective code distance*.

First, let us consider a case without LEC, i.e., error with p_{amb} introduced per qubit per round until the final recovery failure after multiple rounds is evaluated by performing a perfect global decoding. In this case, the error rate is accumulated approximately p_{amb} every round, and the average memory lifetime T can be approximated as

$$T \sim \frac{p_{\text{th}}}{p_{\text{amb}}}, \quad (2)$$

where p_{th} refers to a threshold error rate for the final decoding success evaluation step. Note that p_{th} depends on a type of code: 2D toric code has $p_{\text{th}} \approx 12\%$, the 2D Ising model has $p_{\text{th}} = 50\%$, and 4D toric code has $p_{\text{th}} \approx 4.5\%$. This behavior

TABLE I. Number of data qubit errors to compose an effectively uncorrectable error pattern by entire LEC circuit vs a logical error in each code. L denotes a system size, and D denotes some constant that depends on the LEC circuit.

	LEC's uncorrectable error	Logical error
2D toric code	D	L
2D Ising model	$D \times L$	$L^2/2$
4D toric code	$D \times L$	L^2

is because, for a large enough system size, the final recovery step can be approximated to perform successful decoding if the system's error rate is below p_{th} and failed decoding otherwise.

Now, let us consider a case with LEC: we repeat error generation with rate p_{amb} followed by an LEC circuit application until the decoding failure is evaluated. First, suppose that gates are perfect, i.e., $p_{\text{gate}} = 0$. Since an LEC circuit reduces error density in the system, the LEC circuit is expected to suppress the accumulated rate of errors p_{acc} in each LEC round:

$$p_{\text{acc}} \sim (p_{\text{amb}})^{D_{\text{eff}}}, \quad (3)$$

where the power-law exponent is called *effective code distance* D_{eff} of the LEC circuit. Combining Eqs. (2) and (3), we obtain

$$T \sim \frac{p_{\text{th}}}{p_{\text{acc}}} \sim (p_{\text{amb}})^{-D_{\text{eff}}}. \quad (4)$$

Thus, for the same final recovery step and p_{amb} each LEC round, T gets extended as D_{eff} increases. Since different LEC circuits suppress error accumulation by different amounts, D_{eff} characterizes the performance of the LEC circuit in extending the memory lifetime.

In particular, D_{eff} depends on the size of the *uncorrectable error configuration* by the given LEC circuit. Without LEC, every error in the system survives, gets passed to the next round and contributes to the accumulation of error density. However, with LEC, primarily the “uncorrectable” error configuration by the entire LEC circuit survives and contributes to the error-density accumulation. We summarize the scale of uncorrectable error by LEC compared with logical error in Table I.

Although the above discussion is for LEC circuits with perfect gates, we also define D_{eff} for faulty LEC circuits as well (i.e., $p_{\text{gate}} \neq 0$). Note that the LEC circuit can be interpreted as a dynamical process. The initial error configuration is newly introduced to the system by ambient or gate errors. Then, the LEC circuit is applied to correct such an error pattern. When gates are perfect, the decoding is performed only to these initial errors as shown in Fig. 8(a). However, when gates are faulty, new errors introduced after each parallel gate operation can cause the growth of initial error configuration during the decoding process as shown in Fig. 8(b). Due to this effect, D_{eff} gets decreased when LEC circuits consist of faulty gates. Also, these faulty gates impact the failure rate in each round as well due to this insertion of errors by gate errors during the local decoding process. Thus, we model the average memory

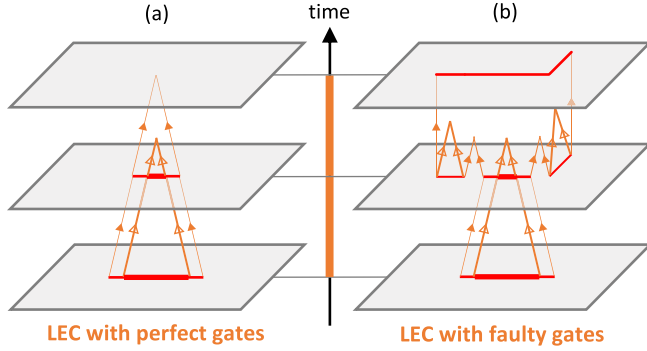


FIG. 8. Effect of faulty gates on effective code distance by comparing a correction of lower-weight (thick red line) vs higher-weight (thin red line) initial error configurations. Local decoding with the same LEC circuit over time is schematically shown for perfect vs faulty gates. (a) When gates are perfect, the LEC circuit removes both of the initial error configurations at the end. (b) When gates are faulty, the LEC circuit still removes an error configuration with a lower weight at the end by the local decoding process (thick orange lines with empty arrows). However, an error configuration with a higher weight “grows” over time by merging with gate errors introduced during the local decoding process (thin orange lines with filled arrows). Thus, we expect the effective code distance for a given LEC circuit to decrease as the circuit’s gate error rate increases.

lifetime T for a given LEC circuit with faulty gates as

$$T \approx C \left(\frac{1}{\alpha \times p_{\text{amb}}} \right)^{D_{\text{eff}}}. \quad (5)$$

Both D_{eff} and α depend on the choice of error-correcting code, p_{gate} , and an LEC circuit. Physically, the parameter D_{eff} describes how fast such errors *grow* over time, whereas the parameter α describes how many new errors (including gate errors) *get introduced* every LEC round. From Table I, we expect that (1) D_{eff} is independent of L for 2D toric code and (2) $D_{\text{eff}} \propto L$ for the 2D Ising model and 4D toric code. Our characterization of memory lifetime by a local decoder is different from the previous literature [38], which follows the quadratic relation between lifetime and ambient error rate. This is due to the difference between gate error regimes of interest.

B. Fitting effective code distance

In Eq. (5), the parameters C , α , and D_{eff} can be determined by fitting the data between average memory lifetime T_L vs ambient error rate p_{amb} , as shown in Fig. 9. To achieve this fitting, we benchmark the average memory lifetime T_L in each error-correcting code with varying input parameters. Then, we compare the RL LEC trained with different p_{gate} and conventional LEC.

In 2D toric code, T is approximately independent of L (see Appendix E2 for a caveat). Thus, from Eq. (5), we fit the following model between T and p_{amb} :

$$\log_{10}(T) = -D_{\text{eff}} \times \log_{10}(p_{\text{amb}}) + k, \quad (6)$$

where $k = \log_{10} C - D_{\text{eff}} \times \log_{10} \alpha$. In the 2D Ising model and 4D toric code, D_{eff} is expected to be proportional to L . Thus, from Eq. (5), we fit the following model between T_L

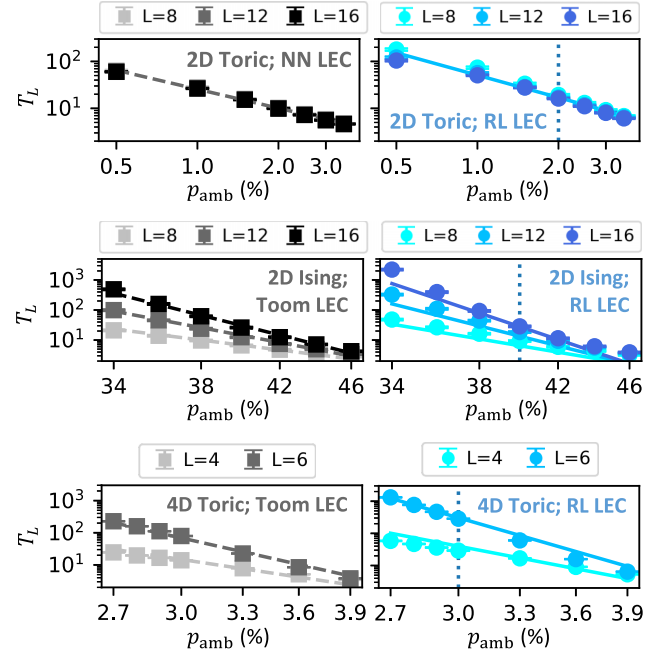


FIG. 9. Fitting from log-log plot of average memory lifetime T_L for each system size L vs ambient error rate p_{amb} for fixed benchmark gate error rate p_{gate} to compute effective code distance. In each code, we perform a simulation of multiple LEC cycles on 1000 independent samples to benchmark T_L , which is the number of LEC cycles until the final perfect decoding results in a logical failure (see Appendix D for the error bar). We compare RL-optimized LEC circuit (blue gradation for varying L) vs nearest-neighbor LEC circuit (NN LEC) in 2D toric code or Toom’s rule-based LEC circuit (Toom LEC) in the 2D Ising model and in 4D toric code (gray gradation for varying L). We train the RL model with $p_{\text{gate}} = 1 \times 10^{-3}$, 1×10^{-4} , and 1×10^{-5} for 2D Ising model and 2D and 4D toric code, respectively. Note that the vertical blue dashed line corresponds to training p_{amb} for each code.

(average memory lifetime obtained by simulating a code with system size L) and p_{amb} :

$$\log_{10}(T_L) = -\frac{D_{\text{eff}}}{L} \times L \times [\log_{10}(p_{\text{amb}}) + k_1] + k_2, \quad (7)$$

where k_1 and k_2 are $\log_{10} \alpha$ and $\log_{10} C$, respectively.

With a fixed LEC circuit, L , and benchmark p_{gate} , let each dataset be the pairs of benchmark p_{amb} and average lifetime. Then, to perform a global fitting based on Eqs. (6) and (7), we use the SymFit Python package [96]. Parameters (D_{eff}, k) for the 2D toric code and (D_{eff}, k_1, k_2) for the 2D Ising model and 4D toric code were shared by datasets varying L with the same LEC circuit and benchmark p_{gate} .

After the fitting, we plot D_{eff} of each LEC circuit vs benchmark p_{gate} , as shown in Fig. 10 (see Appendix E3 for k_1 fitting). Let $p_{\text{gate},c}$ for RL LEC be a critical gate error rate such that RL LEC has higher D_{eff} compared with NN LEC (if 2D toric code) or Toom LEC (if the 2D Ising model or 4D toric code). In every error-correcting code, we observe that the RL LEC trained with lower p_{gate} shows higher fitted D_{eff} in small p_{gate} regime but smaller p_c compared with the LEC circuit trained with higher p_{gate} . For the LEC circuits that are trained with small p_{gate} , as shown in Fig. 10, we additionally

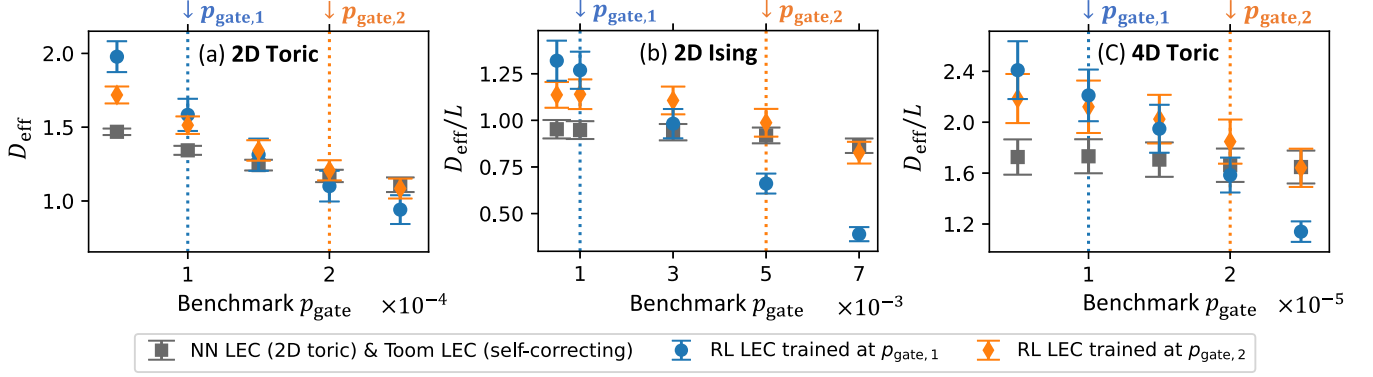


FIG. 10. Fitted effective code distance D_{eff} for given LEC circuits vs benchmark gate error rate p_{gate} in each code (see Appendix D for the error bar). Gray squares represent data with (a) nearest-neighbor LEC circuit (NN LEC) in 2D toric code and (b), (c) Toom’s rule-based LEC circuit (Toom LEC) in the 2D Ising model and 4D toric code. Blue circles (or orange diamond) represent data with RL-optimized LEC circuit trained at p_{gate} with value (a) 1×10^{-4} (or 2×10^{-4}) in 2D toric code, (b) 1×10^{-3} (or 5×10^{-3}) in the 2D Ising model, and (c) 1×10^{-5} (or 2×10^{-5}) in 4D toric code.

find that $p_{\text{gate},c}$ is larger than the training p_{gate} . Note that this property does not apply to LEC circuits trained with large p_{gate} (see Appendix E 4). Also, in the 2D Ising model, the LEC circuit only with Toom actions shows fitted $D_{\text{eff}}/L \approx 1.0$. This is because the minimum uncorrectable error configuration by the Toom LEC circuit consists of L number of data qubit errors as shown in Fig. 4(b). Note that this observation and its explanation are consistent with the intuition from Table I.

VI. APPLICATION OF LOCAL ERROR CORRECTION SCHEME

In this section, we provide two specific examples of how this optimized LEC scheme could be used in various quantum information processing tasks. Section VIA explores how RL-optimized LEC circuits can be used to reduce the number of mid-circuit readouts required to preserve a quantum memory in 2D toric code. On the other hand, Sec. VIB discusses the dissipative preparation of topological phases with a given Hamiltonian.

A. Reducing rate of mid-circuit readouts

Section V shows that LEC circuits can be used to remove errors from encoded qubits and to extend the memory lifetime. Then, it is reasonable to ask whether such LEC circuits can be beneficial to extend memory lifetime in practice. We focus on a 2D toric code to investigate the practical application. In 2D toric code, using only LEC has limitations for a practical application, because memory lifetime extended by LEC does not scale with L as observed in Fig. 9. Also, undecoded error chains get accumulated by the repetition of LECs, which implies that extending memory lifetime has some upper limit. However, we can envision a decoding strategy combining the advantages of both LEC and global decoding. In particular, repeated application of LEC can be used to reduce the rate of error accumulation of an encoded logical qubit. Then, we can periodically perform the global decoding. The key advantage of this approach is that LEC circuits can be applied more

quickly compared with the global decoding that involves mid-circuit readouts.

In practice, we aim for the memory to exhibit a logical error rate P_L below some target value after a given physical time T where each unit time step introduces some error with rate p_{unit} to data qubits. We compare original and hybrid decoding protocols as shown in Fig. 11(a). The original decoding protocol performs $N_{\text{Global}}^{(\text{Original})}$ global decoding with equally spaced time intervals throughout the total time T . The hybrid decoding protocol performs $N_{\text{Global}}^{(\text{Hybrid})}$ global decoding with RL-optimized LEC circuits interleaved between global

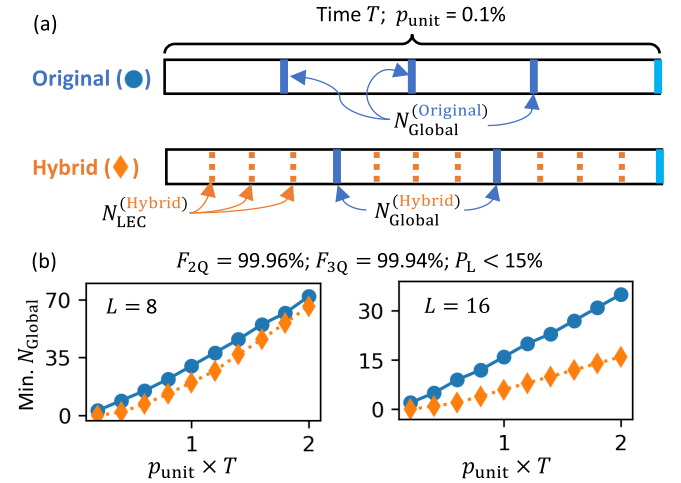


FIG. 11. Original vs hybrid decoding protocol to preserve the 2D toric code memory. (a) At each unit time, new errors with rate $p_{\text{unit}} = 0.1\%$ are introduced to data qubits. Throughout total physical time T , global decoding (blue, solid lines) is performed regularly for both the original (blue circle) and hybrid protocol (orange diamond). However, the hybrid decoding protocol regularly applies LEC circuit (orange, dotted lines) between global decoding applications. After T passes, we evaluate the final logical error rate P_L . (b) Minimum number of global decoding vs $p_{\text{unit}}T$ in each protocol for given multi-qubit gate fidelities and P_L (see Appendix F for details).

decoding. Our goal is to find the regime of parameters, where

$$N_{\text{Global}}^{(\text{Original})} > N_{\text{Global}}^{(\text{Hybrid})}. \quad (8)$$

When two-qubit gate fidelity $F_{2Q} = 99.96\%$, three-qubit gate fidelity $F_{3Q} = 99.88\%$, and target $P_L < 15\%$, two protocols contrast as shown in Fig. 11(b). For an $L = 8$ lattice, the hybrid decoding can reduce the number of global decoding for the constant amount regardless of T . In other words, the *gain* of using the hybrid decoding protocol gets decreased as we want to preserve the memory for a longer time. However, for an $L = 16$ lattice, we observe that the hybrid decoding can reduce the number of global decoding by about half regardless of how long we want to preserve the memory. Note that each global decoding of 2D toric code requires L repetitions of mid-circuit readout of ancillas for faulty syndrome extraction before the 3D minimum-weight perfect matching algorithm [9,10,16]. Thus, we can conclude that using an LEC circuit reduces the number of mid-circuit readouts by $L \times (N_{\text{Global}}^{(\text{Original})} - N_{\text{Global}}^{(\text{Hybrid})})$ throughout the entire physical time, i.e., we can replace some portion of the mid-circuit readouts into possibly more cost-efficient operations. Our investigation confirms that there exists a regime of high multi-qubit gate fidelity where LEC can be practically useful for a memory lifetime enhancement.

B. Dissipatively preparing topological phases

The same methods, developed here for suppressing errors in QEC codes, can be used for the systematic preparation of certain topological phases. The logical subspace of topological error-correcting codes, such as the toric code, is a ground state of the corresponding commuting-projector Hamiltonian [79]. Such states can be prepared in constant time by measuring the stabilizers and then applying an appropriate correction to return to the ground state.

Alternatively, these ground states can be prepared by dissipation, with the circuit depth scaling as the linear size of the system [97]. The procedure described in this work can be interpreted as conditional transport of excitations, resulting in their eventual annihilation and a clean final state. This stabilization (correction) procedure can be used as state preparation when applied to an appropriate initial state.

Consider an example of a 2D toric code, starting in the symmetric product state. This state is stabilized by the X checks but the Z checks have zero expectation value. This can be interpreted as a condensate of the m anyons, which need to be annihilated to prepare the target state. Performing the conditional gates from our protocol amounts to inducing directional transport of these excitations, which eventually leads to their removal and lowering of state energy. Our framework can be used to systematically construct and optimize such dissipative cooling protocols.

Extensions of these methods to more general error-correcting codes, whose parent Hamiltonians can exhibit exceedingly complicated low-energy manifolds [98], is an interesting direction for future work.

VII. DISCUSSION AND OUTLOOK

In this work, we develop an RL framework to optimize measurement-free error-correcting circuits, composed of a

finite number of faulty multi-qubit gates. We specifically focus on finite-sized 2D toric code, the 2D Ising model, and 4D toric code. By introducing a notion of *effective code distance* D_{eff} , we find that when gate fidelities are high enough, the RL-optimized LEC circuits are better than conventional LEC circuits in extending the lifetime of a memory encoded in these codes. Then, we additionally investigate two potential applications of this LEC scheme. First, we show that the portion of global decoding, which consists of mid-circuit readout and classical decoding, can be replaced with LEC circuits with high-fidelity multi-qubit gates for preserving 2D toric code memory. Also, we discuss that quantum states with topological phases can be prepared dissipatively.

Given an error-correcting code and gate error rate, our RL framework learns to optimize the decoding strategy as a sequence composed of operations from a restricted input gate set. Since our task is to optimize such sequences for finite-sized systems, we find RL to be a more suitable tool compared with an analytical method discussed in other literature targeting systems in a thermodynamical limit [81,99].

Our RL framework is adaptive to the needs of various changes. For example, keeping the same gate operations as an input action space, we can simply substitute a biased, realistic error model into our current noise channel. Since coherent errors turn into Pauli errors when we trace out the ancillas at the ancilla reset step [100–103], our RL framework can be used for correcting coherent physical errors in the code as well. Moreover, we can extend our action space by introducing additional error-removal operations that can correct higher-weight errors. It is also possible to apply our RL framework to different codes not discussed in this work as well. For example, it would be interesting to consider higher-dimensional color codes, which support transversal non-Clifford logical operations [104–106], have interesting single-shot error correction properties [107,108], and can also be realized in state-of-the-art quantum platforms [6,109].

Although our RL framework benefits from a simple and small neural network in terms of the training cost, we can choose a more complicated optimizer—including transformer [110] that has already been used in designing an efficient *global* decoder for 2D surface code [22,26]. The advantage of using these more advanced schemes is that we may not need to provide a fixed gate set—depending on our knowledge of error patterns on a given error-correcting code—as input. We can allow RL to find the proper gate input set *from scratch* by only providing the type of available gates as input ingredients. This future direction can also be useful to obtain a *general* LEC circuit optimizer with the stabilizer structure of the code as an input to the RL agent such that we do not have to retrain the agent for each code.

ACKNOWLEDGMENTS

We thank Aurélien Dersy, Yanke Song, Lucas Janson, Timothy H. Hsieh, Tsung-Cheng Lu, Shengqi Sang, Hong-Ye Hu, Di Luo, Andrew K. Saydjari, Kevin B. Xu, Minhak Song, Sangheon Lee, Jin Ming Koh, and Juan Pablo Bonilla Ataides for insightful discussions. We also thank Aurélien Dersy for a critical reading of the paper. We utilize the Stable-Baselines3 package for implementing the RL PPO

algorithm [95], the PyMatching package for implementing the MWPM algorithm for 2D toric code [14,16], and the symfit package for fitting multiple datasets [96]. Our visualization of 2D toric code and the 2D Ising model is motivated by the codebase of previous works [57–59]. M.P. acknowledges funding from Herchel Smith-Harvard Undergraduate Science Research Program (Summer 2022), Harvard Quantum Initiative Undergraduate Research Fellowship (Summer 2022), and Harvard College Research Program (Fall 2022). N.M. acknowledges support from the Department of Energy Computational Science Graduate Fellowship under award number DE-SC0021110. We acknowledge the financial support by the Center for Ultracold Atoms (an NSF Physics Frontiers Center), Wellcome LEAP, Department of Energy/LBNL QSA (Grant No. DE-AC02-05CH11231), NSF (Grant No. PHY-2012023), DARPA ONISQ (Grant No. W911NF2010021), DARPA IMPAQT (Grant No. HR0011-23-3-0030), DARPA MeasQuIT (Grant No. HR0011-24-9-0359), and IARPA ELQ (Grant No. W911NF2320219).

APPENDIX A: EXPERIMENTAL IMPLEMENTATION OF LOCAL ERROR CORRECTION

One of the possible limitations of the LEC circuit is that we regularly reset the ancillas to initial states $|0\rangle$ before we perform a new syndrome extraction operation. In certain state-of-the-art quantum devices, the ancillas with errors can be reset to the initialized ones—regardless of their quantum states—through ancilla repumping techniques [34,37]. Thus, employing these techniques, it is possible for us to perform multiple LEC cycles with a finite number of qubits. Note that the advantage of using the LEC scheme over the standard error correction scheme with ancilla readout is available only if the local coherent operations are more efficient compared with the mid-circuit readout. Thus, an adaptation of this LEC scheme in the actual quantum device should require a careful analysis of resources available within certain quantum platforms.

Moreover, a layout of which gates are applied to which qubits is specific to each syndrome extraction and error-removal operation. Since qubits must be adjacent to each other to perform multi-qubit gates, we change the spatial configuration of data qubits and ancillas before applying each gate for a syndrome extraction and error-removal operation. Note that the qubits can be moved coherently and independently in several state-of-the-art quantum processors, including neutral-atom array platforms [111–113]. Thus, each operation consists of a qubit reconfiguration operation followed by corresponding multi-qubit gates. For simplicity, we ignore the errors from the process of this dynamical rearrangement of qubits and suppose that the qubit reconfiguration operation is perfect.

APPENDIX B: CIRCUIT-LEVEL SIMULATION OF QUANTUM ERROR CORRECTION

In this section, we discuss how we implement the circuit-level simulation of QEC in each code: 2D toric code, the 2D (classical) Ising model, and 4D toric code. Note that for 2D toric code and the 2D Ising model, our visualization of the

lattice, data qubits, and syndromes as in Figs. 2 and 4 are modified based on the codebase of previous works [57,59].

We perform stabilizer-level Monte Carlo simulation based on NumPy for each code, and our error model includes unbiased bit-flip and dephasing errors:

(1) We represent data qubits as three-dimensional arrays: first dimension for a type of a Pauli error on the qubit, second dimension for an index of a copy of a code, and third dimension for an index of a qubit within a code.

(2) To perform a parallelized CNOT, CCX, or CCZ gate operation, we define lists of qubit indices that match with the gate connectivity. The i th index for each list corresponds to control or target qubits for the i th multi-qubit gate to be applied. Although we implement the Pauli error propagation for each CNOT gate, we model each ancilla classically, i.e., either $|0\rangle$ or $|1\rangle$, to avoid dealing with error propagation through non-Clifford CCZ and CCX gates. Physically, this condition can be approximately achieved by resetting ancillas regularly as illustrated in Fig. 1.

We define each parallelized $d = N$ gate or Toom sweep gate as a tessellation of the unit cell as illustrated in Fig. 5. For 2D toric code and 2D Ising model, the visualization is straightforward, as shown in Fig. 3. For 2D toric code, we include four $d = 1$, eight $d = 2$, and twelve $d = 3$ operations with different unit cells as the candidates of error-removal operations. For the 2D Ising model, we include four Toom sweep operations (with 90° rotation), four $d = 1$ operations, and four $d = 2$ operations as the candidates of error-removal operations.

However, for 4D toric code, we effectively represent the data qubits and ancillas using odd and even coordinates [114]. Let us consider a coordinate (x_0, x_1, x_2, x_3) on 4D lattice, where each x_i has an integer value between 0 and $2L - 1$. We denote odd coordinates as o and even coordinates as e to classify qubits:

- (1) Data qubits are on (o, o, e, e) , (o, e, o, e) , (o, e, e, o) , (e, o, o, e) , (e, o, e, o) , and (e, e, o, o) — $6L^2$ in total.
- (2) \hat{Z} stabilizer ancillas are on (o, e, e, e) , (e, o, e, e) , (e, e, o, e) , and (e, e, e, o) — $4L^2$ in total.
- (3) \hat{X} stabilizer ancillas are on (e, o, o, o) , (o, e, o, o) , (o, o, e, o) , and (o, o, o, e) — $4L^2$ in total.

From this representation, we can easily identify six independent logical \hat{Z} and \hat{X} operators. For example, a topologically nontrivial “sheet” of \hat{Z} operators on all data qubits with coordinates $(1, 1, 2k_1, 2k_2)$ with $0 \leq k_1, k_2 \leq L - 1$ correspond to logical \hat{Z} operator, which indeed commutes with all stabilizers.

Now, we introduce error-removal operations based on CCX and CCZ gates in parallel for 4D toric code. As shown in Table II, there are six different configurations of parallelized Toom sweep operations, where each consists of L^2 CCX gates and L^2 CCZ gates. Since each configuration has four choices of “direction” just as in the 2D Ising model illustrated in Fig. 5(b), we have $6 \times 4 = 24$ parallelized Toom sweep operations. On the other hand, CCX and CCZ gates for both $d = 1$ and $d = (1, 1)$ operations are controlled by the same parity classification of ancillas. As shown in Table III, there are four different configurations of parallelized $d = 1$ and $d = (1, 1)$ operations, where each consists of $3L^2/2$ CCX gates and $3L^2/2$ CCZ gates—this is why we choose L to be even. For $d = 1$

TABLE II. Different configurations of parallelized Toom sweep operations. Note that each operation consists of L^2 number of CCX gates and L^2 number of CCZ gates. Also, each configuration has four different “directions.” For example, a CCX gate targeted to data qubit on (1,1,2,2) can be controlled by \hat{Z} stabilizer ancilla on either $(1 \pm 1, 1, 2, 2)$ and either $(1, 1 \pm 1, 2, 2)$.

	Control 1 of CCX	Control 2 of CCX	Target of CCX	Control 1 of CCZ	Control 2 of CCZ	Target of CCZ
Type 1	(o, e, e, e)	(e, o, e, e)	(o, o, e, e)	(e, o, o, o)	(o, e, o, o)	(e, e, o, o)
Type 2	(o, e, e, e)	(e, e, o, e)	(o, e, o, e)	(e, o, o, o)	(o, o, e, o)	(e, o, e, o)
Type 3	(o, e, e, e)	(e, e, e, o)	(o, e, e, o)	(e, o, o, o)	(o, o, o, e)	(e, o, o, e)
Type 4	(e, o, e, e)	(e, e, o, e)	(e, o, o, e)	(o, e, o, o)	(o, o, e, o)	(o, e, e, o)
Type 5	(e, o, e, e)	(e, e, e, o)	(e, o, e, o)	(o, e, o, o)	(o, o, o, e)	(o, e, o, e)
Type 6	(e, e, o, e)	(e, e, e, o)	(e, e, o, o)	(o, o, e, o)	(o, o, o, e)	(o, o, e, e)

operation, each configuration has two choices for a group of data qubits. Thus, we have $4 \times 2 = 8$ parallelized $d = 1$ operations. For $d = (1, 1)$ operation, there are four additional degrees of freedom on top of two choices for a group of data qubits. Thus, we have $4 \times 2 \times 4 = 32$ parallelized $d = (1, 1)$ operations. In summary, we include twenty-four Toom sweep operations, eight $d = 1$ operations, and thirty-two $d = (1, 1)$ operations as the candidates of error-removal operations. For simplifying the RL training and reducing the number of syndrome extraction operations, we “grouped” parallelized $d = 1$ and $d = (1, 1)$ operations each into single actions.

From such design of error-removal operations, we optimize the order of applying parallelized CNOT gates for syndrome extraction operation. For 2D toric code, there are four “steps” of parallelized CNOT gates in sequence to perform a syndrome extraction as illustrated in Fig. 3. For the 2D Ising model and 4D toric code, we perform a syndrome excitation before each error-removal operation. We only update syndromes for the ancillas that would be used in the following error-removal operation. After finding the optimal N number of steps of parallelized CNOT gates in sequence, we test $N!$ possible permutations of this sequence and choose the one with the smallest average number of errors introduced during the entire syndrome extraction operation. Note that it is also possible to optimize such sequence more systematically, as discussed in previous works [115,116].

TABLE III. Different configurations of parallelized $d = 1$ and $d = (1, 1)$ operations. Note that each operation consists of $3L^2/2$ number of CCX gates and $3L^2/2$ number of CCZ gates, and control ancillas are on the same parity classifications of coordinates.

	CCX (Ctrls-Targ)	CCZ (Ctrls-Targ)
Type 1	$(o, e, e, e)-(o, o, e, e)$ $(e, o, e, e)-(e, o, o, e)$ $(e, e, e, o)-(o, e, e, o)$ $(e, o, e, e)-(o, o, e, e)$	$(o, e, o, o)-(e, e, o, o)$ $(o, o, o, e)-(o, e, o, e)$ $(e, o, o, o)-(e, o, e, o)$ $(e, o, o, o)-(e, e, o, o)$
Type 2	$(e, e, o, e)-(e, o, o, e)$ $(o, e, e, e)-(o, e, e, o)$ $(e, e, o, e)-(e, e, o, o)$ $(e, e, o, e)-(e, o, e, o)$	$(o, e, o, o)-(o, e, o, e)$ $(o, o, e, o)-(e, o, e, o)$ $(o, o, e, o)-(o, o, e, o)$ $(o, o, e, o)-(o, o, e, o)$
Type 3	$(e, e, e, o)-(e, o, e, e)$ $(o, e, e, e)-(o, e, o, e)$ $(e, e, e, o)-(e, e, o, o)$ $(e, e, e, o)-(e, o, e, o)$	$(o, e, o, o)-(o, e, e, o)$ $(o, o, o, e)-(e, o, e, o)$ $(o, o, o, e)-(o, o, e, o)$ $(o, o, o, e)-(o, o, e, o)$
Type 4	$(e, o, e, e)-(e, o, e, o)$ $(e, e, o, e)-(o, e, o, e)$	$(o, o, e, o)-(o, e, e, o)$ $(e, o, o, o)-(e, o, e, o)$

APPENDIX C: DETAILS ON REINFORCEMENT LEARNING TRAINING

1. Details on agent

We summarize the RL agent’s construction of an LEC circuit as in Algorithm 2, which defines the “BuildLEC” function in Algorithm 1.

The size of the action space $|\mathcal{A}|$ depends on the code: $|\mathcal{A}| = 24$ for 2D toric code, $|\mathcal{A}| = 12$ for the 2D Ising model, and $|\mathcal{A}| = 26$ for 4D toric code (see Appendix B). Also, in practice, we use two fully connected hidden layers with 128 nodes each for both policy and value networks. We provide a justification of not using bigger networks in Appendix C4.

To represent the previous actions as an observation for the RL agent, we one-hot encoded the LEC action sequence into $|\mathcal{A}|$ -by- H matrix with entries 0 or 1. For 2D toric code, each action A_i is either syndrome extraction or one of error-removal operations in Fig. 5(a), and we fix that the first action A_1 is always a syndrome extraction operation. For the 2D Ising model and 4D toric code, each action A_i is one of the error-removal operations in Fig. 5(b) followed by syndrome extraction operation on corresponding ancillas. We do not fix the first action for the 2D Ising model and 4D toric code, because a syndrome extraction operation is performed for any action before an error-removal operation.

As discussed in Sec. IV B, we construct two different types of RL agents: (1) fixed-depth agent for the 2D Ising model and 4D toric code and (2) variable-depth agent for 2D toric code. For a fixed-depth agent, we consider H to be the fixed LEC circuit depth. However, for a variable-depth agent, we modify the above scheme such that it optimizes the circuit depth $H \leq H_{\max}$, where H_{\max} is a fixed input parameter. To achieve this, we additionally include a *null* action \emptyset where $A_h = \emptyset$ implies that the h th LEC action is “skipped” within a

ALGORITHM 2. “BuildLEC” (RL agent).

Data: Neural network with parameters $f^{\vec{\theta}}$; circuit depth H ; action space \mathcal{A}

Function BuildLEC($f^{\vec{\theta}}, H, \mathcal{A}$):

```

    Circuit  $\leftarrow$  EmptyList;
    for  $h < H$  do
        NewAction :=  $f^{\vec{\theta}}$ (Circuit);
        Circuit.insert(NewAction)
    return Circuit;
```

ALGORITHM 3. “ComputeReward” (RL env.).

Data: Code type T ; lattice size L ; ambient error rate p_{amb} ; gate error rate p_{gate} ; sample size N

Function `ComputeReward`(*circuit*, T , L , p_{amb} , p_{gate} , N):

```

result ← EmptyList;
for sample ∈ {1, ..., N} do
  data ← initialized data qubits;
  for round ∈ {1, ..., R} do
    data := GenerateError(data,  $p_{\text{amb}}$ ,  $T$ );
    ancillas ← initialized ancillas;
    for gate ∈ circuit do
      data, ancillas := ApplyGate(data, ancillas, gate);
      {data, ancillas} := GenerateError({data, ancillas},  $p_{\text{gate}}$ ,  $T$ );
    RewardEach = EvaluateRewardEach(data,  $T$ );
    result.insert(RewardEach);
  return Average of result;

```

Function `GenerateError`(*qubits*, p_{err}):

```

for qubit ∈ qubits do
  randX ∈ Uniform([0,1]);
  if randX <  $p_{\text{err}}$  then
    Apply  $\hat{X}$  error on qubit;
  if  $T$  is 2D or 4D toric code then
    randZ ∈ Uniform([0,1]);
    if randZ <  $p_{\text{err}}$  then
      Apply  $\hat{Z}$  error on qubit;
return qubits;

```

Function `EvaluateRewardEach`(*data*, T):

```

RewardEach ← 0;
if  $T$  is 2D toric code then
  Perform MWPM algorithm on the precise eigenvalues of all stabilizers;
  if Resulting state is a trivial ground state of 2D toric code without logical error then
    RewardEach := 1
if  $T$  is 2D Ising model then
  RewardEach := 1 - (Number of flipped spins among data qubits) /  $L^2$ 
if  $T$  is 4D toric code then
  Perform  $\geq 50$  repetitions of perfect Toom's rule decoding;
  if Resulting state is a trivial ground state of 4D toric code without logical error then
    RewardEach := 1
return RewardEach;

```

sequence of H_{max} actions. Then, removing all \emptyset s from the RL-optimized LEC circuit with H_{max} results in the RL-optimized LEC circuit with H_{opt} as desired.

Two limitations of this modified RL framework are that (1) $H_{\text{opt}} \leq H_{\text{max}}$ and (2) H_{opt} found by RL training is suboptimal if true H_{opt} is much smaller than H_{max} . We set initial H_{max} to be 40 for 2D toric code as we confirmed from training with the fixed-depth agent that H_{opt} is smaller than 40. Then, after the termination condition is satisfied and H_{opt} is obtained, we set the next H_{max} to be the H_{opt} of the previous training and run the next training. This is why we emphasize in Sec. IV B that the training with a variable-depth agent is more “expensive” than the training with a fixed-depth agent.

We have tried an alternative approach to optimize the LEC circuit with a variable depth by adding a *stop* action that terminates the LEC circuit immediately—instead of the \emptyset action. This approach was not successful, because the RL training resulted in building the NN LEC circuit with circuit depth 5 [46], even for training $p_{\text{gate}} = 1 \times 10^{-4}$ where a true

$H_{\text{opt}} > 30$. Since the NN LEC circuit is good enough to possibly form a “deep” local minimum in the reward landscape, we believe that a *stop* action is incentivized enough to be called early. Although it could be possible to get around this issue by modifying the reward structure, we find that the \emptyset -action option is natural to achieve the variable-depth LEC circuit optimization.

2. Details on environment

We summarize the RL environment’s computation of reward as in Algorithm 3, which defines the “ComputeReward” function in Algorithm 1.

We first run a simulation that applies multiple rounds of LEC. In each round, errors are introduced to data qubits, and then the input LEC circuit (fixed across all rounds) is applied to data qubits and newly initialized ancillas. Note that we are interested in extending the lifetime of memory encoded in the error-correcting code, and local decoding is performed

TABLE IV. Code and error model parameters used for optimizing the LEC circuit with our RL framework. L denotes the system size, p_{amb} denotes the ambient error rate in each LEC cycle before applying the LEC circuit, N denotes the number of copies of code, and R denotes the number of repeated LEC cycles used to compute decoding success rate, as illustrated in Fig. 6.

	L	p_{amb}	N	R
2D toric code	8	0.02	100	5
2D Ising model	8	0.40	100	1
4D toric code	4	0.03	50	2

multiple times through the lifetime of memory. Thus, we aim to reduce the *accumulated* error density over multiple rounds. Note that the RL-optimized circuits used in Figs. 7–11 are trained with the code and error parameters in Table IV— p_{gate} varies for each training.

After multiple LEC rounds, we evaluate the decoding success. This *final recovery* step is to determine whether the final state of data qubits after multiple LEC rounds is “close enough” to the trivial ground state of the error-correcting code. This step varies among the choice of the code. For 2D toric code, we perform global decoding by minimum-weight perfect matching (MWPM) classical algorithm with precise syndrome information using PyMatching package [14,16]. For 2D Ising model, we perform a majority vote among data qubits. For 4D toric code, we perform enough (50 rounds) repetition of perfect Toom’s rule decoding. After this step, 2D and 4D toric code decoding are evaluated to be successful if we get a trivial ground state, whereas 2D Ising model decoding is evaluated to be successful if we get less than 50% of all spins bit-flipped. We repeat this simulation N times with a fixed LEC circuit. For each simulation, a stochastic error model results in different configurations of errors applied to qubits. Thus, we can obtain the decoding success rate through this benchmark simulation.

As discussed in Sec. III B, the uncorrectable error configurations after perfect Toom’s rule decoding in 4D toric code can have a low error density. In principle, these uncorrectable errors should be decoded by some nonlocal decoding protocol [44]. However, we decide to leave these uncorrectable errors as decoding failures, because (1) the nonlocal decoding protocol takes significant running time and (2) the prior LEC reduces the uncorrectable error configurations with low error density.

For 2D and 4D toric code, the reward is computed as the number of successful decoding cases divided by the number of total cases N . However, for the 2D Ising model, the reward is computed as an average nonerror density, i.e., average number of nonflipped spins divided by the total number of spins L^2 .

In particular, we choose R in Table IV to be as small as possible for each code to minimize the time for reward computation. However, for 2D toric code, we observe that $R \leq 4$ causes the occurrence of repeating syndrome extraction, which introduces more gate errors into the system without benefiting decoding. This is because, for such a small R , the RL agent is not penalized enough by a bad reward to avoid repeating syndrome extraction and to choose the “do-nothing”

(\emptyset) action instead. Thus, we choose $R = 5$ for the 2D toric code case.

3. Conditions on training termination

Each epoch of our RL training consists of 500 episodes with a minibatch size of 50. In other words, we collect 500 samples of LEC circuits and their benchmarked decoding success rates, divide them into smaller minibatch subsets with 50 samples each, and perform 10 updates with these subsets on the policy and value network. Before saving our updated neural network model, we run 40 epochs for 2D toric code, 80 epochs for the 2D Ising model, and 40 epochs for 4D toric code. We terminate each RL training when a neural network model produces the same LEC circuit as output as the previously saved one throughout these epochs. For 2D toric code with $H_{\text{max}} = 40$ and $p_{\text{gate}} = 1 \times 10^{-4}$, each RL training requires 200–400 epochs. For self-correcting memories with $H = 60$ and any p_{gate} , each RL training requires 800–1000 epochs for the 2D Ising model and 400–600 epochs for 4D toric code. Note that these are obtained with a fully connected neural network with two hidden layers of size 128. Also, we use a variable depth agent for 2D toric code and a fixed depth agent for the 2D Ising model and 4D toric code (see Appendix C 1).

4. Details on reinforcement learning hyperparameters

We utilize the initial given values of hyperparameters for the PPO algorithm in the Stable-Baselines3 library [95] with the following exceptions:

- (1) Since we compute the reward only at the end of each episode, we set a discount factor γ to be one.
- (2) The parameters `n_steps` and `batch_size` are set to be $500H$ and $50H$, where H is H_{max} for 2D toric code and a fixed circuit depth for the 2D Ising model and 4D toric code.
- (3) We use a fully connected neural network with two hidden layers of size 128 for both policy and value networks. The reasoning for this choice is based on the following paragraph.

As shown in Table V, we compare the final reward from RL training with a fully connected neural network with two hidden layers of size d_{NN} with four independent runs. We observe that the final reward gets improved until $d_{\text{NN}} = 128$. However, for $d_{\text{NN}} > 128$, we observe that a similar reward was obtained with shorter training epochs before termination. Considering the size of saved neural networks after training, we choose $d_{\text{NN}} = 128$ for our RL training. Also, this is why we did not have to implement a more complicated neural network structure—such as a transformer—because a simple fully connected neural network with two hidden layers of size 128 was enough for our optimization task.

We attempt the optimization of other hyperparameters—such as `learning_rate` or `n_epochs`—via Optuna framework [117]. However, we find that our RL training has an expensive reward computation and thus is not ideal for such hyperparameter optimization. Furthermore, we observe that, even with the hyperparameters given by Stable Baselines 3, we obtain a satisfactory optimization of the LEC circuits as discussed in Secs. IV and V. Therefore, we did not further investigate such optimization of RL hyperparameters other than the network parameters.

TABLE V. RL training results for 2D toric code and the 2D Ising model with various network sizes d_{NN} . We construct the RL agent with a fully connected neural network with two hidden layers of size d_{NN} (for both the policy and value network of a PPO algorithm) and the RL environment with the parameters shown in the table. In each cell, we record a final reward and the required number of epochs before training termination—“Reward (Number of epochs)” —for each d_{NN} and training index.

2D toric code					2D Ising model				
$(L, H_{\text{max}}N, R, p_{\text{amb}}, p_{\text{gate}}) = (8, 40, 100, 7, 0.02, 1 \times 10^{-4})$					$(L, H_{\text{max}}N, p_{\text{amb}}, p_{\text{gate}}) = (8, 60, 100, 0.40, 1 \times 10^{-3})$				
d_{NN}	Run 1	Run 2	Run 3	Run 4	d_{NN}	Run 1	Run 2	Run 3	Run 4
32	0.1440 (440)	0.1668 (480)	0.1463 (520)	0.1467 (520)	32	0.1022 (1520)	0.1011 (1600)	0.0973 (1360)	0.0981 (1520)
64	0.1668 (280)	0.1468 (360)	0.1489 (400)	0.1634 (320)	64	0.1015 (1120)	0.0975 (1360)	0.1021 (1120)	0.0986 (1120)
128	0.1476 (280)	0.1505 (240)	0.1498 (240)	0.1263 (280)	128	0.0981 (1120)	0.1080 (1040)	0.0973 (1280)	0.0968 (1120)
256	0.1392 (200)	0.1453 (240)	0.1374 (160)	0.1513 (160)	256	0.0964 (960)	0.0991 (800)	0.1017 (720)	0.0972 (880)
512	0.1559 (120)	0.1694 (160)	0.1311 (120)	0.1429 (160)	512	0.0993 (960)	0.1007 (1040)	0.1056 (1120)	0.0953 (1040)

5. Stability of reinforcement learning training

As discussed in Sec. IV A, we choose one LEC circuit out of four RL executions that maximizes the reward. We record rewards by RL LECs from four runs as well as the reward by the conventional LEC with the same parameters as in Table VI. Also, we perform 40 runs of RL training for 2D toric code and the 2D Ising model with $p_{\text{gate}} = 1 \times 10^{-4}$ and 1×10^{-3} , respectively. As shown in Fig. 12, we record the maximum and minimum rewards by RL LECs out of the first 4 and all 40 runs for each code as well as the reward by the conventional LEC with the same parameters. The final reward of any RL LEC is higher than that of conventional LEC, against which we benchmark the decoding performance. Also, we find that the best final reward out of 4 and 40 runs is marginally different, which justifies the number of RL training executions. These results support the stability of RL training, i.e., RL training results do not vary significantly in performance with respect to our demand.

This stability of RL training can be explained by a *simple* state transition, which, in principle, results in low variance in training. In general, RL is applied to problems where the next state s' is obtained from the conditional probability distribution P , given the current state s and action a , i.e., $s' \sim P(s'|s, a)$ [52,83]. For example, in the Gym Acrobot game [118], s' would be updated from s and a based on stochastic laws of physics, which are hidden from the RL agent and have to be “learned.” However, in our problem, s' is determined simply by appending a to s . Thus, our RL training is expected to find the optimized solution more reliably compared with the general application of RL methods, which often report substantial training instabilities [119,120].

TABLE VI. Reward of LEC circuit optimized by four RL training executions and a conventional LEC circuit. The reward is computed out of 10 000 independent samples, and the unit of reward is percent. Also, we choose $p_{\text{gate}} = 1 \times 10^{-4}$ and $p_{\text{gate}} = 1 \times 10^{-5}$ for 2D and 4D toric code, respectively, and $p_{\text{gate}} = 1 \times 10^{-3}$ for the 2D Ising model (with other parameters as in Table IV). For all training executions in the same code, we use the same code and error model parameters.

	RL LEC, Run 1	RL LEC, Run 2	RL LEC, Run 3	RL LEC, Run 4	Conventional LEC
2D Toric Code	90.62 \pm 0.57	90.29 \pm 0.58	91.46 \pm 0.55	91.54 \pm 0.55	86.39 \pm 0.67
2D Ising Model	90.28 \pm 0.58	90.64 \pm 0.57	89.85 \pm 0.59	89.80 \pm 0.59	86.48 \pm 0.67
4D Toric Code	95.28 \pm 0.42	92.82 \pm 0.51	92.47 \pm 0.52	93.90 \pm 0.47	87.06 \pm 0.66

APPENDIX D: ERROR BAR COMPUTATION WITH DATA

Each error bar represents a 95% confidence interval for each parameter.

(1) Final reward (in Figs. 7 and 12):

The final reward is either a decoding success rate (for 2D and 4D toric code) or a “survival” rate of an initial spin direction (for the 2D Ising model). From large- N samples noted in the captions, we compute each of their estimates as a mean of a binomial distribution.

(2) Average memory lifetime T (in Fig. 9):

We compute its estimate as a mean of 1000 samples of independently simulated memory lifetime.

(3) Effective code distance D_{eff} (in Fig. 10):

We obtain fitting parameters and their standard deviations with the SymFit library [96]. Note that we set the parameter `absolute_sigma` to be False.

APPENDIX E: DETAILS ON MEMORY LIFETIME EXTENSION

1. Effect of local error correction on error statistics

In Sec. IV B, we confirm that the RL LEC performs better than the conventional LEC in each code. More specifically, RL LEC shows a lower decoding failure rate compared with conventional LEC after a certain number of repeated LEC cycles. To analyze the reason for this behavior, we classify remaining error patterns after repeated LEC cycle(s) in each code, as shown in Fig. 13. Although both RL and Toom LEC show similar occurrences of logical errors, we observe that RL LEC is effective in suppressing nonlogical errors that are

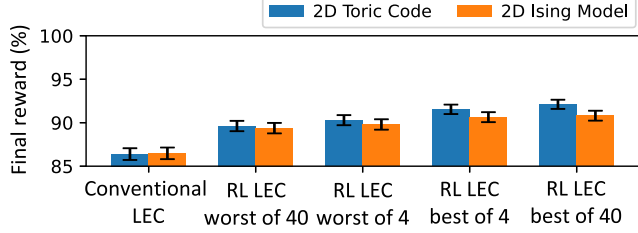


FIG. 12. Final reward comparison of conventional vs RL LEC chosen from worst or best runs out of multiple training executions. Note that conventional LEC is a nearest-neighbor LEC circuit for 2D toric code and Toom’s rule-based LEC circuit for the 2D Ising model and reward is computed out of 10 000 independent samples (see Appendix D for the error bar). RL training for 2D toric code and the 2D Ising model are done with $p_{\text{gate}} = 1 \times 10^{-4}$ and 1×10^{-3} , respectively, and with other parameters as in Table IV.

either correctable or uncorrectable by a single LEC action that consists of the circuits.

In each error-correcting code of Fig. 13, it is indeed expected that the RL LEC reduces the occurrence of the error patterns correctable by their own individual operations. However, over repeated cycles of application, the LEC circuits also reduce the accumulation of higher-weight error patterns that are uncorrectable by their own individual operations. Since these high-weight errors lead to the logical error rate after the

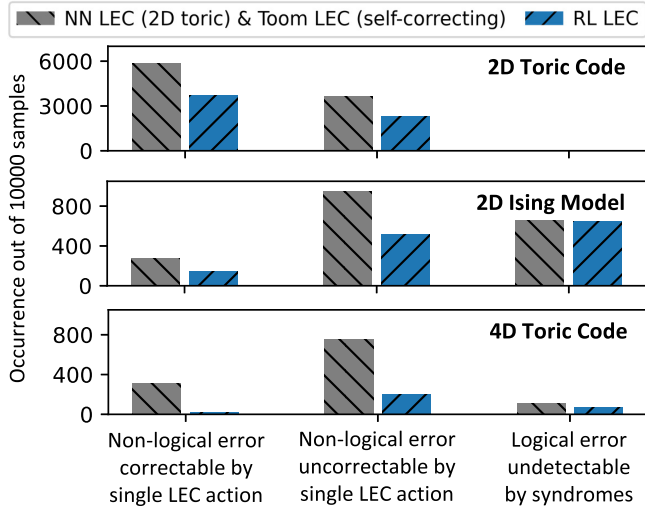


FIG. 13. Classification of remaining errors after LEC cycle(s) in each code. Let a single LEC action be any LEC action among extended error-removal operations as in Fig. 5. Then, we perform a benchmark simulation for applying LEC cycle(s) with 10 000 independent samples. Note that, for each LEC circuit, p_{gate} is 1×10^{-4} for 2D toric code, 1×10^{-3} for the 2D Ising model, and 1×10^{-5} for 4D toric code. Each error pattern after LEC cycle(s) is either (1) a nonlogical *low-weight* error pattern correctable by one of single LEC actions, (2) a nonlogical *high-weight* error pattern uncorrectable by any single LEC action, or (3) a logical error *undetectable* by extracting syndromes. In each code, we compare the occurrence of each category between RL-optimized LEC circuits (blue) trained with parameters in Table IV and above p_{gate} vs conventional LEC circuits (gray).

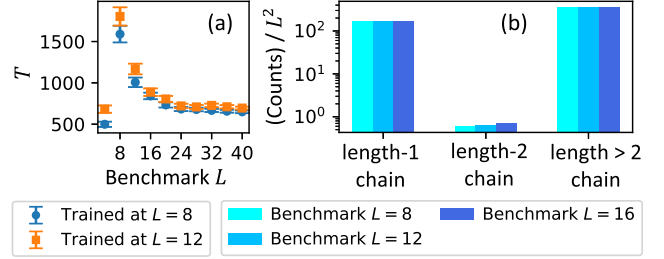


FIG. 14. Comparison between data with varying L benchmarked with perfect gates. (a) Average memory lifetime T vs benchmark lattice size L for an LEC circuit trained with $L = 8$ (blue circles) and an LEC circuit trained with $L = 12$ (orange squares). See (2) of Appendix D for the error bar. For both training, $H_{\text{max}} = 40$, $p_{\text{amb}} = 0.02$, $N = 100$, $R = 7$, and $p_{\text{gate,train}} = 1 \times 10^{-4}$. (b) Number of error chains out of one million samples divided by L^2 for each category of error chain after two LEC rounds. For each benchmark L case, we use the common RL-optimized LEC circuit trained at $L = 8$, $H_{\text{max}} = 40$, $p_{\text{amb}} = 0.02$, $N = 100$, $R = 7$, and $p_{\text{gate,train}} = 1 \times 10^{-4}$. Note that logical error occurs in only three samples for the $L = 8$ case and none for other L s.

final recovery step, *preventing* such errors enable the RL LEC to suppress the logical error rate more effectively compared with the conventional one.

2. Lifetime vs size for two-dimensional toric code

In Fig. 9, we observe that T slightly decreases as L increases for the RL-optimized LEC circuit. We conclude that this is because length- N error chains contribute differently to T . As L gets larger, final recovery by perfect syndrome extraction with perfect ancilla readout and MWPM classical algorithm can correct longer error chains. Thus, the effect of LEC circuits in extending the memory lifetime gets *weaker* due to the varying power of final global decoding as L varies.

Figure 14 supports this explanation. In Fig. 14(a), we find that T decreases as the benchmark L increases for both RL-optimized LEC circuits with perfect gates trained at $L = 8$ and $L = 12$, respectively. Especially, we observe the “drop” of lifetime between benchmark $L = 8$ and $L = 12$. This implies that the relation between a lifetime and lattice size is independent of the training lattice size for optimizing the LEC circuits. Also, in Fig. 14(b), we observe that each length- N error chains occur with a similar *density* regardless of benchmark L for a fixed LEC circuit with perfect gates. This implies that before the final recovery, the effect of the LEC circuit in reducing the number of error chains is similar regardless of benchmark L . Thus, we explain this relation between a lifetime and lattice size based on the role of final recovery.

3. Other parameters from log-log fitting between lifetime and ambient error rate

Although Sec. VB focuses on the fitting parameter D_{eff} from fitting $\log_{10} T$ and $\log_{10}(p_{\text{amb}})$, there exist other fitting parameters as well: k for 2D toric code and k_1 and k_2 for self-correcting memories. For the 2D Ising model and 4D toric code, we plot the fitted k_1 vs benchmark p_{gate} as shown in Fig. 15 to check whether its behavior is consistent with the

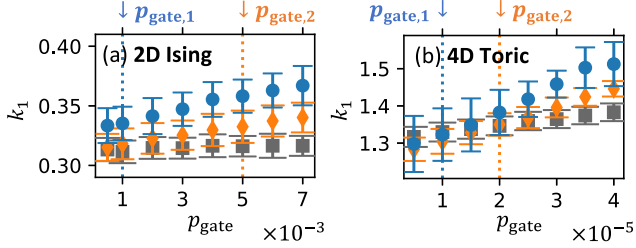


FIG. 15. k_1 vs Benchmark p_{gate} for RL-optimized LEC circuits as in Fig. 10 trained at $p_{\text{gate},1}$ (blue circles) and $p_{\text{gate},2}$ (orange diamonds) as well as Toom LEC circuit (gray square) in (a) a 2D Ising model and (b) 4D toric code. See (3) of Appendix D for the error bar (same as D_{eff}).

intuition on α . Note that we also confirm from the fitting that the standard deviation for k and k_2 do not diverge as well.

Note that k_1 shows different values for different p_{gate} and LEC circuits. We expect that α varies for different p_{gate} and LEC circuits. From Eq. (7), increasing k_1 implies increasing α , because a logarithm function is monotonically increasing. Physically, α denotes how many new errors *get introduced* every LEC round. Then, we can explain two general trends observed in Fig. 15 based on this intuition on α .

For each LEC circuit, k_1 increases as benchmark p_{gate} increases. This trend is consistent with the intuition on α —as p_{gate} increases, more new errors are introduced to the system. Also, for each benchmark p_{gate} , k_1 is the biggest for the RL-optimized LEC circuit trained with smaller p_{gate} and the smallest for the Toom LEC circuit. Note that, in Sec. IV B, we show that RL-optimized LEC circuit trained with larger p_{gate} includes less higher-weight error-removal operations. This is because the higher-weight error-removal operations may introduce more errors into the system compared with the Toom operation. Thus, for the same benchmark p_{gate} , k_1 is largest for the most “complicated” LEC circuits that are trained with smaller p_{gate} and smallest for the “simplest” Toom circuits.

4. Local error correction circuits with large gate error rate

For 2D toric code, RL training with large p_{gate} results in the simplest NN decoder. Note that we fix our first action in the LEC circuit to be a syndrome extraction (see Appendix C 1). Thus, we do not observe that the RL framework produces a do-nothing LEC circuit. However, we find that for $p_{\text{gate}} > 0.01$, the simplest NN circuit performs worse than just do-nothing.

On the other hand, as shown in Fig. 16, we observe that the RL-optimized LEC circuit trained at $p_{\text{gate},\text{train}}$ performs as much as—or even worse than—the Toom LEC circuit when benchmark p_{gate} is $p_{\text{gate},\text{train}}$ if $p_{\text{gate},\text{train}}$ is large enough. In other words, we observe that $p_{\text{gate},c} \approx p_{\text{gate},\text{train}}$ or even $p_{\text{gate},c} < p_{\text{gate},\text{train}}$ for large $p_{\text{gate},\text{train}}$ case, unlike as shown in Fig. 10. This feature is the limitation of our RL framework, and this is the reason why we do not investigate improving the threshold gate error rate of Toom’s rule decoder using our RL framework.

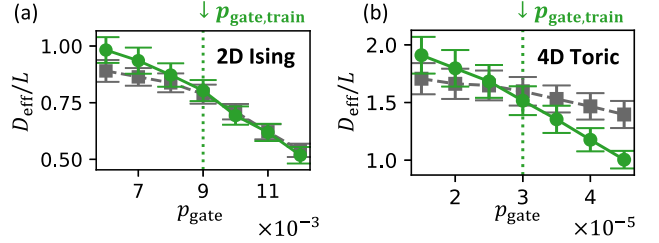


FIG. 16. D_{eff} vs Benchmark p_{gate} for RL-optimized LEC circuit trained at large $p_{\text{gate},\text{train}}$ (green circles) as well as Toom LEC circuit (gray squares) in (a) a 2D Ising model and (b) 4D toric code. See (3) of Appendix D for the error bar. Here, we choose $p_{\text{gate},\text{train}}$ to be 9×10^{-3} for the 2D Ising model and 3×10^{-5} for 4D toric code and other parameters to be as in Table IV.

APPENDIX F: DETAILS ON LOCAL ERROR CORRECTION APPLICATION

From our choice of the error model, the fidelity of a CNOT gate $F_{2Q} = 1 - 4 \times p_{\text{gate}}$, and the fidelity of a CCX or CCZ gate $F_{3Q} = 1 - 6 \times p_{\text{gate}}$. Thus, multi-qubit gates with $p_{\text{gate}} = 1 \times 10^{-4}$ have $F_{2Q} = 99.96\%$ and $F_{3Q} = 99.94\%$. In this regime, we investigate what is the minimum number of global MWPM decoding for the original vs hybrid decoding protocol to achieve the final logical error rate after total time T to be below 15%.

As shown in Fig. 17, we perform 1D and 2D optimization in Sec. VIA for the original and hybrid decoding protocol, respectively. For the original decoding, we vary N_{MWPM} and check the final P_L after time T . On the other hand, for the hybrid decoding, we first fix a total number of global decoding and then find N_{LEC} that minimizes P_L after time T . Then, we choose the minimum number of global MWPM decoding for each of the protocols that satisfy $P_L < 15\%$.

On top of the minimum N_{MWPM} for $p_{\text{gate}} = 1 \times 10^{-4}$ discussed in Sec. VIA, we also compare minimum N_{MWPM} for larger p_{gate} as well. When $p_{\text{gate}} = 3 \times 10^{-3}$, i.e., $F_{2Q} = 99.88\%$ and $F_{3Q} = 99.82\%$, we obtain that the hybrid protocol can reduce the number of global decoding by one or two compared with the original protocol regardless of T . Also, note that we observe the same results for these comparisons, even though the original decoding protocol is improved by replacing the 3D MWPM following repeated faulty syndrome extractions with the 2D MWPM following perfect syndrome extraction.

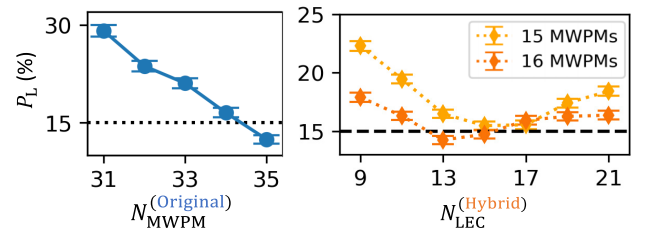


FIG. 17. Method of finding minimum number of global decoding for original vs hybrid decoding protocol in Sec. VIA to achieve $P_L < 15\%$ for fixed $T = 2000$ with 5000 samples of 2D toric code lattice with $L = 16$. See (1) of Appendix D for the error bar (same as final reward).

- [1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000), pp. 425–427.
- [2] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [3] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of real-time fault-tolerant quantum error correction, *Phys. Rev. X* **11**, 041058 (2021).
- [4] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton *et al.* Suppressing quantum errors by scaling a surface code logical qubit, *Nature (London)* **614**, 676 (2023).
- [5] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, V. Vuletić, and M. D. Lukin, A quantum processor based on coherent transport of entangled atom arrays, *Nature (London)* **604**, 451 (2022).
- [6] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, Logical quantum processor based on reconfigurable atom arrays, *Nature (London)* **626**, 58 (2024).
- [7] D. Gottesman, Class of quantum error-correcting codes saturating the quantum hamming bound, *Phys. Rev. A* **54**, 1862 (1996).
- [8] D. Gottesman, Stabilizer codes and quantum error correction, [arXiv:quant-ph/9705052](https://arxiv.org/abs/quant-ph/9705052).
- [9] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
- [10] 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).
- [11] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [12] N. Delfosse, Decoding color codes by projection onto surface codes, *Phys. Rev. A* **89**, 012317 (2014).
- [13] N. H. Nickerson and B. J. Brown, Analysing correlated noise on the surface code using adaptive decoding algorithms, *Quantum* **3**, 131 (2019).
- [14] O. Higgott, Pymatching: A python package for decoding quantum codes with minimum-weight perfect matching, [arXiv:2105.13082](https://arxiv.org/abs/2105.13082).
- [15] A. M. iOlius, J. E. Martinez, P. Fuentes, and P. M. Crespo, Performance enhancement of surface codes via recursive minimum-weight perfect-match decoding, *Phys. Rev. A* **108**, 022401 (2023).
- [16] O. Higgott and C. Gidney, Sparse blossom: correcting a million errors per core second with minimum-weight matching, [arXiv:2303.15933](https://arxiv.org/abs/2303.15933).
- [17] N. Maskara, A. Kubica, and T. Jochym-O'Connor, Advantages of versatile neural-network decoding for topological codes, *Phys. Rev. A* **99**, 052351 (2019).
- [18] C. Chinni, A. Kulkarni, D. M. Pai, K. Mitra, and P. K. Sarvepalli, Neural decoder for topological codes using pseudo-inverse of parity check matrix, [arXiv:1901.07535](https://arxiv.org/abs/1901.07535).
- [19] S. Varsamopoulos, K. Bertels, and C. G. Almudever, Comparing neural network based decoders for the surface code, *IEEE Trans. Comput.* **69**, 300 (2020).
- [20] D. Bhoumik, P. Sen, R. Majumdar, S. Sur-Kolay, L. K. K. J., and S. S. Iyengar, Efficient decoding of surface code syndromes for error correction in quantum computing, [arXiv:2110.10896](https://arxiv.org/abs/2110.10896).
- [21] E. Egorov, R. Bondesan, and M. Welling, The end: An equivariant neural decoder for quantum error correction, [arXiv:2304.07362](https://arxiv.org/abs/2304.07362).
- [22] J. Bausch, A. W. Senior, F. J. H. Heras, T. Edlich, A. Davies, M. Newman, C. Jones, K. Satzinger, M. Y. Niu, S. Blackwell, G. Holland, D. Kafri, J. Atalaya, C. Gidney, D. Hassabis, S. Boixo, H. Neven, and P. Kohli, Learning to decode the surface code with a recurrent, transformer-based neural network, [arXiv:2310.05900](https://arxiv.org/abs/2310.05900).
- [23] S. Gicev, L. C. L. Hollenberg, and M. Usman, A scalable and fast artificial neural network syndrome decoder for surface codes, *Quantum* **7**, 1058 (2023).
- [24] T. Fu and Y. Zhao, Benchmarking machine learning models for quantum error correction, [arXiv:2311.11167](https://arxiv.org/abs/2311.11167).
- [25] B. Hall, S. Gicev, and M. Usman, Artificial neural network syndrome decoding on ibm quantum processors, *Phys. Rev. Res.* **6**, L032004 (2024).
- [26] H. Wang, P. Liu, K. Shao, D. Li, J. Gu, D. Z. Pan, Y. Ding, and S. Han, Transformer-QEC: Quantum error correction code decoding with transferable transformers, [arXiv:2311.16082](https://arxiv.org/abs/2311.16082).
- [27] S. Bravyi, M. Suchara, and A. Vargo, Efficient algorithms for maximum likelihood decoding in the surface code, *Phys. Rev. A* **90**, 032326 (2014).
- [28] B. Criger and I. Ashraf, Multi-path Summation for decoding 2D topological codes, *Quantum* **2**, 102 (2018).
- [29] N. Delfosse and N. H. Nickerson, Almost-linear time decoding algorithm for topological codes, *Quantum* **5**, 595 (2021).
- [30] N. Delfosse and G. Zémor, Linear-time maximum likelihood decoding of surface codes over the quantum erasure channel, *Phys. Rev. Res.* **2**, 033042 (2020).
- [31] W. Xu, A. V. Venkatramani, S. H. Cantú, T. Šumarac, V. Klüsener, M. D. Lukin, and V. Vuletić, Fast preparation and detection of a Rydberg qubit using atomic ensembles, *Phys. Rev. Lett.* **127**, 050501 (2021).
- [32] E. Deist, Y.-H. Lu, J. Ho, M. K. Pasha, J. Zeiher, Z. Yan, and D. M. Stamper-Kurn, Mid-circuit cavity measurement in a neutral atom array, *Phys. Rev. Lett.* **129**, 203602 (2022).
- [33] M. A. Norcia, W. B. Cairncross, K. Barnes, P. Battaglini, A. Brown, M. O. Brown, K. Cassella, C.-A. Chen, R. Cox, D. Crow, J. Epstein, C. Griger, A. M. W. Jones, H. Kim, J. M. Kindem, J. King, S. S. Kondov, K. Kotru, J. Lauigan, M. Li *et al.*, Midcircuit qubit measurement and rearrangement in a ¹⁷¹Yb atomic array, *Phys. Rev. X* **13**, 041034 (2023).
- [34] T. M. Graham, L. Phuttitarn, R. Chinnarasu, Y. Song, C. Poole, K. Jooya, J. Scott, A. Scott, P. Eichler, and M. Saffman, Mid-circuit measurements on a single-species neutral alkali atom quantum processor, *Phys. Rev. X* **13**, 041051 (2023).
- [35] K. Singh, C. E. Bradley, S. Anand, V. Ramesh, R. White, and H. Bernien, Mid-circuit correction of correlated phase errors using an array of spectator qubits, *Science* **380**, 1265 (2023).

- [36] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, H. Levine, G. Semeghini, M. Greiner, V. Vuletić, and M. D. Lukin, High-fidelity parallel entangling gates on a neutral-atom quantum computer, *Nature (London)* **622**, 268 (2023).
- [37] J. W. Lis, A. Senoo, W. F. McGrew, F. Rönchen, A. Jenkins, and A. M. Kaufman, Midcircuit operations using the *omg* architecture in neutral atom arrays, *Phys. Rev. X* **13**, 041035 (2023).
- [38] N. P. Breuckmann, K. Duivenvoorden, D. Michels, and B. M. Terhal, Local decoders for the 2D and 4D toric code, *Quantum Info. Comput.* **17**, 181 (2017).
- [39] M. Herold, E. T. Campbell, J. Eisert, and M. J. Kastoryano, Cellular-automaton decoders for topological quantum memories, *npj Quantum Inf.* **1**, 15010 (2015).
- [40] A. Kubica and J. Preskill, Cellular-automaton decoders with provable thresholds for topological codes, *Phys. Rev. Lett.* **123**, 020501 (2019).
- [41] M. Vasmer, D. E. Browne, and A. Kubica, Cellular automaton decoders for topological quantum codes with noisy measurements and beyond, *Sci. Rep.* **11**, 2027 (2021).
- [42] T. R. Scruby and K. Nemoto, Local probabilistic decoding of a quantum code, *Quantum* **7**, 1093 (2023).
- [43] A. Ray, R. Laflamme, and A. Kubica, Protecting information via probabilistic cellular automata, *Phys. Rev. E* **109**, 044141 (2024).
- [44] N. P. Breuckmann and X. Ni, Scalable neural network decoders for higher dimensional quantum codes, *Quantum* **2**, 68 (2018).
- [45] K. Meinerz, C.-Y. Park, and S. Trebst, Scalable neural decoder for topological surface codes, *Phys. Rev. Lett.* **128**, 080505 (2022).
- [46] S. C. Smith, B. J. Brown, and S. D. Bartlett, Local predecoder to reduce the bandwidth and latency of quantum error correction, *Phys. Rev. Appl.* **19**, 034050 (2023).
- [47] S. Heußen, D. F. Locher, and M. Müller, Measurement-free fault-tolerant quantum error correction in near-term devices, *PRX Quantum* **5**, 010333 (2024).
- [48] T. L. M. Guedes, D. Winter, and M. Müller, Quantum cellular automata for quantum error correction and density classification, *Phys. Rev. Lett.* **133**, 150601 (2024).
- [49] M. A. Perlin, V. N. Premakumar, J. Wang, M. Saffman, and R. Joynt, Fault-tolerant measurement-free quantum error correction with multiqubit gates, *Phys. Rev. A* **108**, 062426 (2023).
- [50] S. Veroni, M. Müller, and G. Giudice, Optimized measurement-free and fault-tolerant quantum error correction for neutral atoms, [arXiv:2404.11663](https://arxiv.org/abs/2404.11663).
- [51] Y. Zhu and M.-J. Hwang, Passive error correction with a qubit-oscillator system in noisy environment, [arXiv:2408.05145](https://arxiv.org/abs/2408.05145).
- [52] L. P. Kaelbling, M. L. Littman, and A. W. Moore, Reinforcement learning: a survey, *J. Artif. Int. Res.* **4**, 237 (1996).
- [53] M. Krenn, J. Landgraf, T. Foesel, and F. Marquardt, Artificial intelligence and machine learning for quantum technologies, *Phys. Rev. A* **107**, 010101 (2023).
- [54] H. P. Nautrup, N. Delfosse, V. Dunjko, H. J. Briegel, and N. Friis, Optimizing quantum error correction codes with reinforcement learning, *Quantum* **3**, 215 (2019).
- [55] V. P. Su, C. Cao, H.-Y. Hu, Y. Yanay, C. Tahan, and B. Swingle, Discovery of optimal quantum error correcting codes via reinforcement learning, [arXiv:2305.06378](https://arxiv.org/abs/2305.06378).
- [56] J. Olle, R. Zen, M. Puviani, and F. Marquardt, Simultaneous discovery of quantum error correction codes and encoders with a noise-aware reinforcement learning agent, [arXiv:2311.04750](https://arxiv.org/abs/2311.04750).
- [57] P. Andreasson, J. Johansson, S. Liljestrand, and M. Granath, Quantum error correction for the toric code using deep reinforcement learning, *Quantum* **3**, 183 (2019).
- [58] R. Sweke, M. S. Kesselring, E. P. L. van Nieuwenburg, and J. Eisert, Reinforcement learning decoders for fault-tolerant quantum computation, *Mach. Learn.: Sci. Technol.* **2**, 025005 (2021).
- [59] D. Fitzek, M. Eliasson, A. F. Kockum, and M. Granath, Deep q-learning decoder for depolarizing noise on the toric code, *Phys. Rev. Res.* **2**, 023230 (2020).
- [60] E. S. Matekole, E. Ye, R. Iyer, and S. Y.-C. Chen, Decoding surface codes with deep reinforcement learning and probabilistic policy reuse, [arXiv:2212.11890](https://arxiv.org/abs/2212.11890).
- [61] Y. J. Patel, A. Kundu, M. Ostaszewski, X. Bonet-Monroig, V. Dunjko, and O. Danaci, Curriculum reinforcement learning for quantum architecture search under hardware errors, [arXiv:2402.03500](https://arxiv.org/abs/2402.03500).
- [62] A. Kundu, Reinforcement learning-assisted quantum architecture search for variational quantum algorithms, [arXiv:2402.13754](https://arxiv.org/abs/2402.13754).
- [63] J. Chen, Z. Cai, K. Xu, D. Wu, and W. Cao, Qubit-wise architecture search method for variational quantum circuits, [arXiv:2403.04268](https://arxiv.org/abs/2403.04268).
- [64] V. N. Ivanova-Rohling, N. Rohling, and G. Burkard, Discovery of an exchange-only gate sequence for cnot with record-low gate time using reinforcement learning, [arXiv:2402.10559](https://arxiv.org/abs/2402.10559).
- [65] O. Menegasso Pires, E. Inacio Duzzioni, J. Marchi, and R. De Santiago, Quantum circuit synthesis using projective simulation, *Intel. Artif.* **24**, 90 (2021).
- [66] S. Z. Baba, N. Yoshioka, Y. Ashida, and T. Sagawa, Deep reinforcement learning for preparation of thermal and prethermal quantum states, *Phys. Rev. Appl.* **19**, 014068 (2023).
- [67] C.-C. Li, R.-H. He, and Z.-M. Wang, Enhanced quantum state preparation via stochastic predictions of neural networks, *Phys. Rev. A* **108**, 052418 (2023).
- [68] M. Kölle, T. Schubert, P. Altmann, M. Zorn, J. Stein, and C. Linnhoff-Popien, A reinforcement learning environment for directed quantum circuit synthesis, [arXiv:2401.07054](https://arxiv.org/abs/2401.07054).
- [69] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement learning with neural networks for quantum feedback, *Phys. Rev. X* **8**, 031084 (2018).
- [70] R. Zen, J. Olle, L. Colmenarez, M. Puviani, M. Müller, and F. Marquardt, Quantum circuit discovery for fault-tolerant logical state preparation with reinforcement learning, [arXiv:2402.17761](https://arxiv.org/abs/2402.17761).
- [71] C. Ryan-Anderson, N. C. Brown, C. H. Baldwin, J. M. Dreiling, C. Foltz, J. P. Gaebler, T. M. Gatterman, N. Hewitt, C. Holliman, C. V. Horst, J. Johansen, D. Lucchetti, T. Mengle, M. Matheny, Y. Matsuoka, K. Mayer, M. Mills, S. A. Moses, B. Neyenhuis, J. Pino, P. Siegfried, R. P. Stutz, J. Walker, and D. Hayes, High-fidelity and fault-tolerant teleportation of a logical qubit using transversal

- gates and lattice surgery on a trapped-ion quantum computer, [arXiv:2404.16728](#).
- [72] Y.-J. Liu and S. Lieu, Dissipative phase transitions and passive error correction, *Phys. Rev. A* **109**, 022422 (2024).
- [73] I. Cong, N. Maskara, M. C. Tran, H. Pichler, G. Semeghini, S. F. Yelin, S. Choi, and M. D. Lukin, Enhancing detection of topological order by local error correction, *Nat. Commun.* **15**, 1527 (2024).
- [74] A. L. Toom, Nonergodic multidimensional system of automata, *Probl. Peredachi Inf.* **10**, 70 (1974).
- [75] A. L. Toom, Stable and attractive trajectories in multicomponent systems, in *Multicomponent Random Systems*, Advances in Probability and Related Topics, edited by R. Dobrushin, I. Sina, and D. Griffeath (Marcel Dekker, New York, 1980), pp. 549–575.
- [76] F. V. Day and S. Barrett, The ising ferromagnet as a self-correcting physical memory: a monte-carlo study, [arXiv:1201.0390](#).
- [77] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, On thermal stability of topological qubit in Kitaev’s 4D model, *Open Sys. Info. Dyn.* **17**, 1 (2010).
- [78] L. F. Gray, Toom’s stability theorem in continuous time, in *Perplexing Problems in Probability: Festschrift in Honor of Harry Kesten*, edited by M. Bramson and R. Durrett (Birkhäuser Boston, Boston, 1999), pp. 331–353.
- [79] A. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys. (NY)* **303**, 2 (2003).
- [80] J. Kazemí and H. Weimer, Genuine bistability in open quantum many-body systems, [arXiv:2111.05352](#).
- [81] 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).
- [82] S. Roberts and S. D. Bartlett, Symmetry-protected self-correcting quantum memories, *Phys. Rev. X* **10**, 031041 (2020).
- [83] D. J. Foster and A. Rakhlin, Foundations of reinforcement learning and interactive decision making, [arXiv:2312.16730](#).
- [84] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, [arXiv:1707.06347](#).
- [85] G. Dulac-Arnold, R. Evans, H. van Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin, Deep reinforcement learning in large discrete action spaces, [arXiv:1512.07679](#).
- [86] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, Robust adversarial reinforcement learning, [arXiv:1703.02702](#).
- [87] I. Popov, N. Heess, T. Lillicrap, R. Hafner, G. Barth-Maron, M. Vecerik, T. Lampe, Y. Tassa, T. Erez, and M. Riedmiller, Data-efficient deep reinforcement learning for dexterous manipulation, [arXiv:1704.03073](#).
- [88] Y.-L. Tuan, J. Zhang, Y. Li, and H.-yi Lee, Proximal policy optimization and its dynamic version for sequence generation, [arXiv:1808.07982](#).
- [89] E. Hansen, A. Barto, and S. Zilberstein, Reinforcement learning for mixed open-loop and closed-loop control, in *Advances in Neural Information Processing Systems*, edited by M. Mozer, M. Jordan, and T. Petsche (MIT Press, 1996), Vol. 9.
- [90] H. Ghraieb, J. Viquerat, A. Larcher, P. Meliga, and E. Hachem, Single-step deep reinforcement learning for open-loop control of laminar and turbulent flows, *Phys. Rev. Fluids* **6**, 053902 (2021).
- [91] A. Raffin, O. Sigaud, J. Kober, A. Albu-Schäffer, J. Silvério, and F. Stulp, An open-loop baseline for reinforcement learning locomotion tasks, [arXiv:2310.05808](#).
- [92] E.-J. Kuo, Y.-L. L. Fang, and S. Y.-C. Chen, Quantum architecture search via deep reinforcement learning, [arXiv:2104.07715](#).
- [93] S. Katoch, S. S. Chauhan, and V. Kumar, A review on genetic algorithm: past, present, and future, *Multimedia Tools Appl.* **80**, 8091 (2021).
- [94] X. Liu, P. Li, F. Meng, H. Zhou, H. Zhong, J. Zhou, L. Mou, and S. Song, Simulated annealing for optimization of graphs and sequences, *Neurocomputing* **465**, 310 (2021).
- [95] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, Stable-baselines3: Reliable reinforcement learning implementations, *J. Mach. Learn. Res.* **22**, 1 (2021).
- [96] M. Roelfs and P. C. Kroon, [symfit](#) (2017), doi:[10.5281/zenodo.1133336](#).
- [97] M. Kalinowski, N. Maskara, and M. D. Lukin, Non-Abelian Floquet spin liquids in a digital Rydberg simulator, *Phys. Rev. X* **13**, 031008 (2023).
- [98] A. Anshu, N. P. Breuckmann, and C. Nirkhe, NLTS Hamiltonians from good quantum codes, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023 (Association for Computing Machinery, New York, 2023), pp. 1090–1096.
- [99] Y. Hong, J. Guo, and A. Lucas, Quantum memory at nonzero temperature in a thermodynamically trivial system, [arXiv:2403.10599](#).
- [100] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, *npj Quantum Inf.* **4**, 55 (2018).
- [101] J. K. Iverson and J. Preskill, Coherence in logical quantum channels, *New J. Phys.* **22**, 073066 (2020).
- [102] Y. Zhao and D. E. Liu, An analytic study of the independent coherent errors in the surface code, [arXiv:2112.00473](#).
- [103] Á. Márton and J. K. Asbóth, Coherent errors and readout errors in the surface code, *Quantum* **7**, 1116 (2023).
- [104] A. Kubica, B. Yoshida, and F. Pastawski, Unfolding the color code, *New J. Phys.* **17**, 083026 (2015).
- [105] J. Haah and M. B. Hastings, Measurement sequences for magic state distillation, *Quantum* **5**, 383 (2021).
- [106] D. Honciuc Menendez, A. Ray, and M. Vasmer, Implementing fault-tolerant non-Clifford gates using the $[[8,3,2]]$ color code, *Phys. Rev. A* **109**, 062438 (2024).
- [107] H. Bombín, Single-shot fault-tolerant quantum error correction, *Phys. Rev. X* **5**, 031043 (2015).
- [108] B. J. Brown, N. H. Nickerson, and D. E. Browne, Fault-tolerant error correction with the gauge color code, *Nat. Commun.* **7**, 12302 (2016).
- [109] M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, A. Chernoguzov, J. M. Dreiling, C. Foltz, F. Frachon, J. P. Gaebler, T. M. Gatterman, L. Grans-Samuelsson, D. Hayes, N. Hewitt, J. Johansen, D. Lucchetti, M. Mills, S. A. Moses, B. Neyenhuis, A. Paz, J. Pino, P. Siegfried *et al.*, Demonstration of logical qubits and repeated error correction with better-than-physical error rates, [arXiv:2404.02280](#).

- [110] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, [arXiv:1706.03762](#).
- [111] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Probing many-body dynamics on a 51-atom quantum simulator, *Nature (London)* **551**, 579 (2017).
- [112] W. Tian, W. J. Wee, A. Qu, B. J. M. Lim, P. R. Datla, V. P. W. Koh, and H. Loh, Parallel assembly of arbitrary defect-free atom arrays with a multitweezer algorithm, *Phys. Rev. Appl.* **19**, 034048 (2023).
- [113] Q. Xu, J. P. Bonilla Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasić, M. D. Lukin, L. Jiang, and H. Zhou, Constant-overhead fault-tolerant quantum computation with reconfigurable atom arrays, *Nat. Phys.* **20**, 1084 (2024).
- [114] T. Jochym-O'Connor and T. J. Yoder, Four-dimensional toric code with non-Clifford transversal gates, *Phys. Rev. Res.* **3**, 013118 (2021).
- [115] T. Tansuwannont, C. Chamberland, and D. Leung, Flag fault-tolerant error correction, measurement, and quantum computation for cyclic Calderbank-Shor-Steane codes, *Phys. Rev. A* **101**, 012342 (2020).
- [116] C. Ryan-Anderson, N. C. Brown, M. S. Allman, B. Arkin, G. Asa-Attuah, C. Baldwin, J. Berg, J. G. Bohnet, S. Braxton, N. Burdick, J. P. Campora, A. Chernoguzov, J. Esposito, B. Evans, D. Francois, J. P. Gaebler, T. M. Gatterman, J. Gerber, K. Gilmore, D. Gresh *et al.*, Implementing fault-tolerant entangling gates on the five-qubit code and the color code, [arXiv:2208.01863](#).
- [117] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, [arXiv:1907.10902](#).
- [118] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, Openai gym, [arXiv:1606.01540](#).
- [119] L. Buşoniu, T. de Bruin, D. Tolić, J. Kober, and I. Palunko, Reinforcement learning for control: Performance, stability, and deep approximators, *Annu. Rev. Control* **46**, 8 (2018).
- [120] R. Wang, Y. Wu, R. Salakhutdinov, and S. M. Kakade, Instabilities of offline RL with pre-trained neural representation, [arXiv:2103.04947](#).