Emergent Unitary Designs for Encoded Qubits from Coherent Errors and Syndrome Measurements

Zihan Cheng,[1,*] Eric Huang,[2] Vedika Khemani,[3] Michael J. Gullans,[2] and Matteo Ippoliti[1]

[1] *Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA*

[2] *Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*

[3] *Department of Physics, Stanford University, Stanford, California 94305, USA*

(Received 26 April 2025; revised 10 July 2025; accepted 24 July 2025; published 22 August 2025)

Unitary k-designs are distributions of unitary gates that match the Haar distribution up to its kth statistical moment. They are a crucial resource for randomized quantum protocols. However, their implementation on encoded logical qubits is nontrivial due to the need for magic gates, which can require a large resource overhead. In this work, we propose an efficient approach to generate unitary designs for encoded qubits in surface codes by applying local unitary rotations ("coherent errors") on the physical qubits followed by syndrome measurement and error correction. We prove that, under some conditions on the coherent errors (notably including all single-qubit unitaries) and on the error-correcting code, this process induces a unitary transformation of the logical subspace. We numerically show that the ensemble of logical unitaries (indexed by the random syndrome outcomes) converges to a unitary design in the thermodynamic limit, provided that the density or strength of coherent errors is above a finite threshold. This "unitary design" phase transition coincides with the code's coherent error threshold under optimal decoding. Furthermore, we propose a classical algorithm to simulate the protocol based on a "staircase" implementation of the surface code encoder and decoder circuits. This enables a mapping to a $(1+1)$-dimensional monitored circuit, where we observe an entanglement phase transition (and thus a classical complexity phase transition of the decoding algorithm) coinciding with the aforementioned unitary design phase transition. Our results provide a practical way to realize unitary designs on encoded qubits, with applications including quantum state tomography and benchmarking in error-correcting codes.

DOI: 10.1103/bnld-2chd

I. INTRODUCTION

Quantum error correction is crucial to the realization of scalable quantum computations with real, noisy hardware. At the same time, encoded quantum information tends to be harder to manipulate. Quantum error-correcting codes have some restricted set of logical operations that can be implemented "transversally," i.e., by separately rotating individual physical qubits—this is desirable for fault tolerance, as transversal operations do not propagate errors. But, these transversal logical operations are not universal for quantum computing [1]. For surface codes and other codes of practical interest, "magic" (i.e., non-Clifford) gates are among those that cannot be implemented transversally. Strategies to overcome this issue, such as magic state distillation, come with some overhead in terms of auxiliary qubits or circuit size [2–5]. This motivates the search for alternative strategies that can implement generic unitary operations on encoded qubits without additional resources.

In this work we address this issue within the context of *random unitary gates*, which are a key component of various quantum information processing protocols like randomized benchmarking, randomized measurements for state and process learning, cryptography, random circuit sampling, quantum simulation, etc. Our strategy is based on the application of transversal physical operations (which may be seen as "coherent errors," though applied intentionally and known to the experimentalist) followed by syndrome extraction and error correction, as sketched in Fig. 1(a). The Born-rule randomness inherent in quantum measurements automatically gives rise to a random distribution of operations on the encoded information; remarkably, under some criteria that we characterize,

*Contact author: zihan_cheng@utexas.edu

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

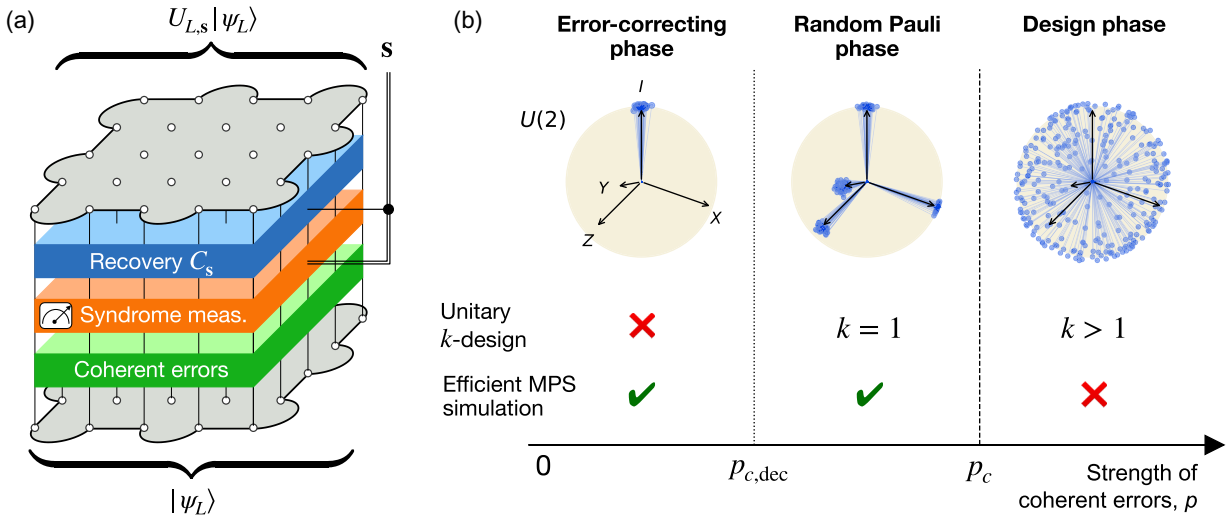


FIG. 1. Main ideas of this work. (a) Intentional application of coherent errors to a quantum error-correcting code, followed by syndrome measurement and correction, yields (under the assumptions of Theorem 1) a logical unitary operation that depends on the syndrome outcome **s**, i.e., a projected ensemble of logical unitaries. (b) Depending on the strength of the coherent errors, the projected ensemble can take sharply different forms in the limit of large code size: an "error-correcting phase," where all unitaries in the ensemble are very close to the identity with high probability; a "unitary design phase," where the distribution of unitaries becomes highly uniform over the logical unitary group $U(2)$; and an intervening "random Pauli phase," where the distribution clusters around Pauli unitaries I, X, Y, Z with equal probability. Qualitative sketches of these behaviors are shown at the top. Threshold $p_{c,\mathrm{dec}}$ between the error-correcting and random-Pauli phases depends on the choice of decoding algorithm, while threshold p_c of the unitary design phase is intrinsic (achieved by the optimal decoder). We find that it also corresponds to a transition in complexity for an MPS-based classical simulation algorithm.

these operations are unitary, and the resulting ensemble can achieve a universal form in the thermodynamic limit (namely, a unitary k-design, i.e., an approximation to the Haar distribution on the unitary group).

This phenomenon is reminiscent of "deep thermalization" [6–9], the emergence of universal distributions of postmeasurement *states* on a spatially local subsystem from measuring its environment, but with two key differences: the distribution is over unitary operators instead of states, and the partition between measured and unmeasured degrees of freedom is not based on real space, but rather on *syndrome* versus *logical* degrees of freedom.

In this work we focus on the two-dimensional surface code, and find that the emergence of unitary designs in the thermodynamic limit occurs only above a critical strength (or density) of the coherent errors, as summarized in Fig. 1(b). Remarkably, this threshold for design formation is also a threshold for three other related but distinct phenomena: (i) an *error-correction* threshold for the code under coherent errors and optimal decoding; (ii) an *entanglement* phase transition in an associated $(1+1)$-dimensional monitored dynamics; (iii) a *computational complexity* phase transition for a classical decoder based on matrix product states that takes advantage of the aforementioned entanglement transition. Our results add to other recently discovered examples of coding [10], "magic" [11], entanglement [12], and delocalization [13] transitions in similar setups involving coherent errors on encoded qubits.

The rest of this paper is structured as follows. In Sec. II we review essential background concepts including state and unitary designs, projected ensembles, and measurement-induced phase transitions. Section III contains our first main result, regarding projected ensembles of unitary operations arising from coherent errors and syndrome measurements in error-correcting codes. In Sec. IV we numerically study the emergence of unitary designs from these ensembles, focusing on the surface code. We then introduce a mapping to a $(1+1)$-dimensional [$(1+1)$D] monitored dynamics via staircase encoder-decoder circuits in Sec. V, where we numerically uncover an entanglement transition and analytically connect it to the unitary design threshold. Finally, we discuss our results and point to future directions in Sec. VI.

II. BACKGROUND

A. State designs and unitary designs

Many protocols in quantum information science take advantage of the properties of random unitary operations or random quantum states [14–18]. However, Haar-random quantum states or unitaries have exponentially high complexity, making them impractical. This motivates the introduction of *designs*—distributions of quantum states or

unitaries that capture only some statistical moments of the Haar distribution, but can be implemented efficiently [19–23]. Here we briefly review the definitions of quantum state and unitary designs and related concepts.

Definition 1 (State ensemble, moment operator). An ensemble of states is a probability distribution over the Hilbert space: $\mathcal{E} = \{p_i, |\psi_i\rangle\}$ with state $|\psi_i\rangle$ occurring with probability p_i. Given an ensemble of states $\mathcal{E} = \{p_i, |\psi_i\rangle\}$, its kth moment operator is

$$\rho_{\mathcal{E}}^{(k)} = \mathbb{E}_{\psi\sim\mathcal{E}}[(|\psi\rangle\langle\psi|)^{\otimes k}] = \sum_i p_i (|\psi_i\rangle\langle\psi_i|)^{\otimes k}. \quad (1)$$

This is a mixed state on k copies of the system.

Definition 2 (Unitary ensemble, twirling channel). An ensemble of unitaries is a probability distribution over the unitary group $U(D)$ (D is the Hilbert space dimension): $\mathcal{E} = \{p_i, U_i\}$ with the unitary operator U_i occurring with probability p_i. Given an ensemble of unitaries $\mathcal{E} = \{p_i, U_i\}$, its k-fold twirling channel is

$$\Phi_{\mathcal{E}}^{(k)}[\cdot] = \mathbb{E}_{U\sim\mathcal{E}}[U^{\otimes k}[\cdot](U^\dagger)^{\otimes k}] = \sum_i p_i U_i^{\otimes k}[\cdot](U_i^\dagger)^{\otimes k}. \quad (2)$$

This is a channel on k copies of the system.

Remark 1. More generally, for both definitions, one could use a continuous rather than atomic probability measure, but that is not necessary for the purpose of this work.

We are interested in ensembles of quantum states or unitaries that approximate certain statistical moments of the Haar distribution. This notion is made precise by the following definitions.

Definition 3 (ϵ-approximate k-design). We say that an ensemble of states $\mathcal{E}$ forms an ϵ-approximate quantum state k-design if

$$\Delta_{\text{state}}^{(k)} = \|\rho_{\mathcal{E}}^{(k)} - \rho_{\text{Haar}}^{(k)}\|_1 \le \epsilon, \quad (3)$$

where $k \ge 1$ is an integer, $\|A\|_1 = \frac{1}{2}\text{Tr}(\sqrt{A^\dagger A})$ is the trace norm, and $\rho_{\text{Haar}}^{(k)} = \mathbb{E}_{\psi\sim\text{Haar}}(|\psi\rangle\langle\psi|)^{\otimes k}$. We say that an ensemble of unitaries $\mathcal{E}$ forms an ε-approximate unitary k-design if

$$\Delta^{(k)} = \|\Phi_{\mathcal{E}}^{(k)} - \Phi_{\text{Haar}}^{(k)}\|_\diamond \le \epsilon, \quad (4)$$

where

$$\|\Phi\|_\diamond = \max_X \frac{\|(\Phi\otimes\mathcal{I})(X)\|_1}{\|X\|_1} \quad (5)$$

is the diamond norm and $\Phi_{\text{Haar}}^{(k)}[\cdot] = \mathbb{E}_{U\sim\text{Haar}} U^{\otimes k}[\cdot](U^\dagger)^{\otimes k}$. Here $\mathcal{I}$ is the identity channel on a second copy of the Hilbert space, and X is chosen among operators acting on both copies.

B. Projected ensembles

The projected ensemble, introduced in Refs. [6,7], is an ensemble of quantum states generated by projective measurements on part of a many-body system. The Born-rule randomness of measurement outcomes induces a probability distribution over wave functions of the subsystem left unmeasured; remarkably, this distribution can be a quantum state k-design in many cases of interest. More generally, the emergence of universal distributions (including but not limited to k-designs) from the projected ensemble goes under the name of "deep thermalization" [6–9,24–27]. Here we briefly review the construction of the projected ensemble for states; our work will generalize this construction to an ensemble of unitaries.

Consider a pure quantum state $|\Psi\rangle_{AB}$ on a bipartite system AB consisting of N_A qubits in A and N_B qubits in B. Upon performing projective measurements on B, one obtains a measurement outcome (bitstring) $\mathbf{z} \in \{0,1\}^{N_B}$ on B and a corresponding postmeasurement state on A, $|\psi_{\mathbf{z}}\rangle_A$. The projected ensemble can be defined as

$$\mathcal{E}_{\psi,A} = \{p(\mathbf{z}), |\psi_{\mathbf{z}}\rangle_A\}, \quad (6)$$

where $|\psi_{\mathbf{z}}\rangle_A = {}_B\langle\mathbf{z}|\Psi\rangle_{AB}/\sqrt{p(\mathbf{z})}$ is the (normalized) postmeasurement state and $p(\mathbf{z})$ is the Born-rule probability of outcome $\mathbf{z} \in \{0,1\}^{N_B}$.

It has been shown that if $|\Psi\rangle_{AB}$ is a random many-body state, for sufficiently large N_B, the projected ensemble $\mathcal{E}_{\psi,A}$ forms an approximate state k-design with high probability [7]. Other universal distributions have been identified for Hamiltonian eigenstates at variable energy [7,26], for fermionic or bosonic Gaussian states [28,29], or for random states that respect a symmetry [30].

It is natural to ask whether a similar construction could be used to realize not just state designs, but also unitary designs. However, a naïve attempt to generalize this construction runs into the issue that the operations generated by a projected ensemble are typically *not* unitary. To be precise, consider a unitary transformation U_{AB} acting on a bipartite state $|\psi\rangle_A \otimes |0\rangle_B$, where $|0\rangle_B$ is some fixed state of B, while $|\psi\rangle_A$ is an arbitrary, variable input state on A; projective measurements on B then give rise to an ensemble of operations on A, $\{K_{\mathbf{z}} = {}_B\langle\mathbf{z}| U_{AB} |0\rangle_B\}$. These are Kraus operators that unravel a quantum channel $\Phi(\cdot) = \text{Tr}_B[U_{AB}([\cdot]_A \otimes |0\rangle\langle 0|_B)U_{AB}^\dagger]$ for which U_{AB} is a Stinespring dilation. Generically, these Kraus operators are nonunitary, but rather correspond to weak measurements of subsystem A. This is also manifested in the fact that the probability distribution over Kraus operators $K_{\mathbf{z}}$ depends on the input state $|\psi\rangle$ on A, $p(\mathbf{z}) = \langle\psi| K_{\mathbf{z}}^\dagger K_{\mathbf{z}} |\psi\rangle$.

So $\mathcal{E}$ is not an intrinsically defined ensemble of transformations. The two issues (unitarity of the transformations and independence of the probability distribution from the input state) are in fact equivalent.

Fact 1. Consider a set of Kraus operators $\{K_{\mathbf{z}}\}$. The probability distribution $p(\mathbf{z}) = \mathrm{Tr}(K_{\mathbf{z}}\rho K_{\mathbf{z}}^\dagger)$ is independent of the input state ρ if and only if each $K_{\mathbf{z}}$ is proportional to a unitary operator:

$$K_{\mathbf{z}} = \sqrt{p(\mathbf{z})}U_{\mathbf{z}}. \tag{7}$$

Proof. This is a well-known fact in quantum information theory; here we report a simple proof for the sake of completeness. The *if* statement is obvious. As for the *only if* statement, we start by defining the positive operator-valued measure $\{E_{\mathbf{z}} = K_{\mathbf{z}}^\dagger K_{\mathbf{z}}\}$, such that $p(\mathbf{z}) = \mathrm{Tr}(E_{\mathbf{z}}\rho)$ for any state ρ. Let $\{|\alpha\rangle\}$ be an orthonormal basis of the Hilbert space; then

$$E_{\mathbf{z},\alpha\alpha} = \mathrm{Tr}(E_{\mathbf{z}}|\alpha\rangle\langle\alpha|) = p(\mathbf{z}) \quad \text{for all } \alpha. \tag{8}$$

Furthermore, for any pair of basis vectors $|\alpha\rangle, |\beta\rangle$ $(\beta \neq \alpha)$, we can let $\rho = \frac{1}{2}(|\alpha\rangle + |\beta\rangle)(\langle\alpha| + \langle\beta|)$, which yields

$$E_{\mathbf{z},\alpha\beta} + E_{\mathbf{z},\beta\alpha} = 0. \tag{9}$$

The same reasoning for $\rho = \frac{1}{2}(|\alpha\rangle + i|\beta\rangle)(\langle\alpha| - i\langle\beta|)$ yields

$$E_{\mathbf{z},\alpha\beta} - E_{\mathbf{z},\beta\alpha} = 0. \tag{10}$$

Combining Eqs. (8)–(10), we have

$$E_{\mathbf{z},\alpha\beta} = p(\mathbf{z})\delta_{\alpha\beta}, \tag{11}$$

and thus $K_{\mathbf{z}} = \sqrt{p(\mathbf{z})}U_{\mathbf{z}}$ for some unitaries $U_{\mathbf{z}}$. ■

Past works have identified a setting where one can obtain projected ensembles of unitary transformations, $\mathcal{E} = \{p_{\mathbf{z}}, U_{\mathbf{z}}\}$, where the $U_{\mathbf{z}}$ are unitary and the $p(\mathbf{z})$ are intrinsically defined; the key ingredient is quantum error correction. Specifically Bravyi *et al.* [31] focused on the surface code and identified certain unitary ensembles (made only of z rotations) that can be realized from syndrome measurements. In this work, we build on that result to allow for the realization of projected ensembles of *general* unitaries on a single qubit, thus enabling the formation of unitary designs.

C. Monitored circuits and the space-evolving block decimation algorithm

Recent years have seen a surge of interest in "monitored" quantum dynamics [32,33]: time evolution where unitary interactions coexist with measurements. The interest stems from the discovery of sharp thresholds, or phase transitions, in the structure of quantum correlations in these dynamics or in their output states. While there are many variations of these phenomena, the simplest instance is represented by local unitary circuits interspersed with projective measurements, where the rate p of measurements (the probability that a qubit is measured after each gate) tunes a transition from an area-law-entangled phase to a volume-law-entangled one [34–37]. These phases are closely related to ideas of quantum error correction and decoding [38–44].

A significant application of monitored quantum dynamics lies in the classical simulation of quantum dynamics—even in the absence of measurements. As an example, noise in the form of single-qubit dephasing or depolarizing channels can be unraveled into measurements; quantum trajectory methods then effectively simulate a monitored time evolution, and it may be possible to leverage the entanglement phase transition for efficient tensor network simulation (depending on the spatial dimensionality) [45–47].

Another way to leverage the entanglement phase transition for efficient classical simulation, with or without noise, applies to sampling experiments on shallow circuits [46,48]. There, an approach dubbed "space-evolving block decimation" (SEBD) effectively trades one spatial dimension for a time dimension, turning the D-dimensional shallow unitary circuit into a deep monitored circuit in $D-1$ dimensions. Loosely speaking, by sweeping through the circuit along one spatial direction, the measurements occurring at the end of the original circuit (responsible for the sampling) are effectively spread out throughout the dynamics. Crudely associating a "measurement rate" $p \sim 1/T$, where T is the shallow circuit's depth (each qubit participates in T unitary gates and one measurement), yields a phase transition in the complexity of SEBD controlled by the depth, with efficient simulation below a finite critical depth T_c.

In this work we adapt the SEBD approach to the distinct setting of syndrome measurements on the surface code—a nontrivial generalization since single-qubit measurements are essential to the formulation of SEBD. To this end, we leverage "staircase" encoder and decoder circuits for surface codes.

III. PROJECTED ENSEMBLE OF LOGICAL UNITARIES

In this section we study conditions under which syndrome measurements on an error-correcting code give rise to unitary operations, building on and extending the results of Bravyi *et al.* [31].

Based on the discussion in Sec. II B, and specifically Fact 1, a projected ensemble of *unitary* operators requires

that the measurements reveal no information about the input state. This naturally points to the framework of quantum error correction, where syndrome measurements are carefully chosen so as to reveal no information about the logical input state. Consider a quantum error-correcting code initialized in a logical state, then subject to some unitary evolution U followed by syndrome measurement and correction. Generally, U takes the system outside the code subspace, while the syndrome measurement and correction returns it to the code subspace. The goal of error correction is to return the system to the same logical state it started from, but depending on U this may fail, so the final logical state is generally different from the initial one. We therefore obtain an ensemble of linear transformations of the code space, and it is natural to ask under what conditions these transformations are unitary.

A setting where this was shown to be the case is the surface code under coherent Z errors [31,49]. Bravyi *et al.* [31] showed that starting with a logical state $|\psi_L\rangle$ of the surface code, single-qubit Z rotations $U = \prod_j \exp(i\eta_j Z_j)$ followed by syndrome measurement and correction result in an effective logical Z rotation $U_{L,\mathbf{s}} = \exp(i\theta_{\mathbf{s}} Z_L)$ (indexed by the syndrome outcome $\mathbf{s}$) provided that the surface code has odd distance. Importantly, by Fact 1, unitarity of $U_{L,\mathbf{s}}$ implies that the syndrome probability distribution $p(\mathbf{s})$ is independent of the initial logical state $|\psi_L\rangle$. As a result, we have a well-defined projected ensemble of logical unitaries:

$$\mathcal{E}_{U,L} = \{p(\mathbf{s}), \exp(i\theta_{\mathbf{s}} Z_L)\}. \tag{12}$$

However, the proof technique employed in Ref. [31] only applies to Z rotations (or, equivalently, X rotations, by the CSS duality of the surface code), which limits the projected ensemble to rotations about a single axis. Such ensembles clearly fail to form unitary designs for the $U(2)$ logical group [they may at most form designs for a $U(1)$ subgroup].

Here we generalize the result above in two ways: by considering codes beyond the surface code, and by allowing physical operations beyond single-qubit Z (or X) rotations. This allows the realization of general $U(2)$ unitaries on the logical qubit.

We consider a general stabilizer code prepared in a logical state $|\psi_L\rangle$, and subject to a physical unitary U:

$$|\psi'\rangle = U|\psi_L\rangle\,. \tag{13}$$

State $|\psi'\rangle$ is generally not a logical state (U may be seen as a "coherent error" on the code). Our goal is to find constraints on U and on the error-correcting code such that the final logical state, obtained from $|\psi'\rangle$ by syndrome measurement and correction, is related to the input logical state $|\psi_L\rangle$ by a logical unitary operation that is *independent of the input state* [see Fig. 1(a)].

To address this question, we must first introduce some notation. Let $\{g_i\}$ be a minimal set of stabilizer generators for the error-correcting code. We introduce the syndrome projectors $\Pi_{\mathbf{s}} = \prod_i [\mathbb{I} + (-1)^{s_i} g_i]/2$. To each syndrome we associate a recovery operation $C_{\mathbf{s}}$, i.e., a Pauli unitary that obeys $\Pi_{\mathbf{s}} C_{\mathbf{s}} = C_{\mathbf{s}} \Pi_{\mathbf{0}}$ (it returns error states with syndrome $\mathbf{s}$ to the code space, characterized by the trivial syndrome $\mathbf{0}$). Note that the $C_{\mathbf{s}}$ are only defined modulo stabilizer and logical operations—the specific choice of recovery $\{C_{\mathbf{s}}\}$ defines a decoder. Our statements here hold for any choice of $\{C_{\mathbf{s}}\}$. With this notation, we can write the final state of the protocol as

$$|\phi_{\mathbf{s},L}\rangle = \frac{C_{\mathbf{s}}\Pi_{\mathbf{s}} U|\psi_L\rangle}{\| C_{\mathbf{s}}\Pi_{\mathbf{s}} U|\psi_L\rangle \|}, \tag{14}$$

occurring with probability $\langle\psi_L| U^\dagger \Pi_{\mathbf{s}} U|\psi_L\rangle$ (the probability of measuring syndrome $\mathbf{s}$). We aim to understand when the above reduces to an ensemble of unitary operations on the logical space. We answer this question by proving the following theorem.

Theorem 1. Consider a stabilizer code with one logical qubit and a unitary U on the physical qubits. Assume that the following three conditions hold.

(1) All the stabilizers have even Pauli weight.
(2) The code is CSS, and its X and Z code distances d_x, d_z are odd.
(3) The physical unitary U commutes with the time-reversal transformation $\mathcal{T} = \mathcal{K}\prod_j (iY_j)$ (with $\mathcal{K}$ the complex conjugation).

Then

$$C_{\mathbf{s}}\Pi_{\mathbf{s}} U\Pi_{\mathbf{0}} = \sqrt{p(\mathbf{s})}\, U_{L,\mathbf{s}}\Pi_{\mathbf{0}}, \tag{15}$$

where $p(\mathbf{s})$ is a probability distribution over syndrome outcomes $\mathbf{s}$ and the $U_{L,\mathbf{s}}$ are unitaries on the logical subspace.

Proof. Here we only provide an outline of the proof; see Appendix A for details. For any Pauli operator P, one has $\mathcal{T}P\mathcal{T}^{-1} = (-1)^{|P|}P$ with $|P|$ the Pauli weight (number of nonidentity Pauli matrices) of P. Conditions (1) and (2) imply that all nontrivial logical operators have odd Pauli weight, so all the logical information is stored in time-reversal-odd operators; this remains true after application of U, by condition (3). But, the syndrome measurements only reveal information about time-reversal-even operators [by condition (1) on the stabilizer generators], so they reveal no information about the logical state. Unitarity then follows by Fact 1. ■

Corollary 1. For any logical input state $|\psi_L\rangle$, the syndrome probability distribution on $U|\psi_L\rangle$ is $p(\mathbf{s}) =$

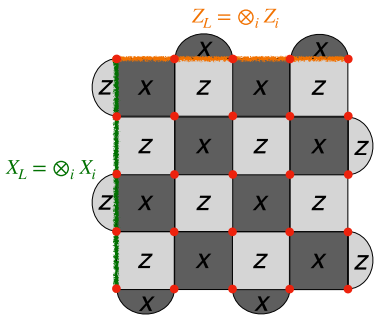


FIG. 2. Rotated surface code with distance $d = 5$. Physical qubits are denoted by red points. Logical operators X_L and Z_L are denoted by green and orange strings of X's and Z's, respectively. We have $|X_L| = |Z_L| = 5$ and $|Y_L| = |iX_LZ_L| = 9$, so all logical operators have odd weight.

$\frac{1}{2}\mathrm{Tr}(U\Pi_{\mathbf{0}}U^\dagger\Pi_{\mathbf{s}})$, independent of the input state; the output state conditional on syndrome $\mathbf{s}$ is

$$C_{\mathbf{s}}\Pi_{\mathbf{s}}U|\psi_L\rangle \propto U_{L,\mathbf{s}}|\psi_L\rangle\,. \tag{16}$$

Theorem 1 asserts that, when conditions (1)–(3) on the error-correcting code and physical operation U are met, one indeed obtains a projected ensemble of unitary operations on the logical subspace. We next discuss when these conditions arise.

Condition (1), on the Pauli weight of stabilizers, is met whenever one has a set of stabilizer generators $\{g_i\}$ where each g_i has even Pauli weight [50]. This is true in many codes of interest, for instance in the toric code (four-body generators), rotated surface code (four-body generators in the bulk and two-body generators on the boundaries; Fig. 2), color code (six-body generators), etc.

Condition (2) is straightforward and can typically be realized in families of codes such as the surface code (Fig. 2) or the color code by a choice of system size. We note that this condition can be relaxed, as our proof only uses a weaker property: that all nontrivial logical operators must have odd Pauli weight. This is true, for instance, in the five-qubit perfect code, which is non-CSS.

Condition (3) is most easily met by setting $U = \bigotimes_{i=1}^{n} u_i$, where $\{u_i\}$ are arbitrary single-qubit unitaries. It is easy to see that $u = e^{-i\theta\mathbf{n}\cdot\boldsymbol{\sigma}/2}$ (rotation by θ around the $\mathbf{n}$ axis) commutes with the time-reversal symmetry, since $\mathcal{T}\sigma^\alpha = -\sigma^\alpha\mathcal{T}$ for $\alpha = x, y, z$, and thus $\mathcal{T}i\mathbf{n}\cdot\boldsymbol{\sigma} = -i\mathcal{T}\mathbf{n}\cdot\boldsymbol{\sigma} = +i\mathbf{n}\cdot\boldsymbol{\sigma}\mathcal{T}$, using the antiunitarity of $\mathcal{T}$. More generally, the condition $[U, \mathcal{T}] = 0$ identifies a subgroup [51] of $U(2^n)$, comprising all unitaries $U = e^{iH}$, where H is a real linear combination of odd-weight Pauli operators. Thus, Theorem 1 covers a large class of *many-body* coherent errors U. As an example, one can consider local circuits of three-body gates such as $e^{-i\alpha ZZZ}$, interspersed by arbitrary single-qubit gates.

We note that the unitarity of the logical operation was proven for the special case of single-qubit Z rotations on the odd-distance surface code in Ref. [31], and recently generalized to arbitrary single-qubit rotations in Ref. [52]. Theorem 1 generalizes the result to a much wider class of codes and coherent errors.

For the rest of the present work, we focus on rotated surface codes with odd distance d, subject to arbitrary single-qubit physical rotations. The generalization to many-body physical unitaries is an interesting direction whose exploration we leave to future work.

IV. EMERGENT LOGICAL UNITARY DESIGNS

Theorem 1 allows the realization of projected ensembles of logical unitary operations

$$\mathcal{E}_{U,L} = \{p(\mathbf{s}), U_{\mathbf{s}}\} \tag{17}$$

that in principle go beyond Z rotations, thus extending the results of Ref. [31] [see Eq. (12)]. In analogy with the phenomenon of "deep thermalization" in projected ensembles of states (Sec. II B), it is then natural to ask about the emergence of universal behavior in this ensemble, and specifically about the possibility of unitary designs.

When U is weak enough, i.e., close enough to identity, the "coherent error" represented by U is correctable, which means that all logical unitaries in the projected ensemble approach the logical identity in the limit of large code distance, $d \to \infty$. Therefore, logical unitary designs are only possible above the error threshold, i.e., for strong enough U. In the following we numerically show that unitary designs emerge as soon as coherent errors exceed the optimal coherent error threshold. The optimal error threshold thus also serves as the critical point for a "unitary design phase transition."

Before proceeding, we first remark on the role of the decoder. Similar to error correction, the projected ensemble in Eq. (17) depends on the choice of decoder, i.e., the assignment of a recovery operation $C_{\mathbf{s}}$ to each syndrome $\mathbf{s} \in \{0,1\}^{n-1}$. With different decoders, the logical unitaries may differ by a logical Pauli unitary for some syndrome outcomes. Specifically, if a decoder yields corrections $\{C_{\mathbf{s}}\}$ and another yields corrections $\{C'_{\mathbf{s}} = \sigma_{\mathbf{s}}C_{\mathbf{s}}\}$ (with $\sigma_{\mathbf{s}}$ logical Pauli unitaries), then the projected ensembles for the two decoders will be

$$\mathcal{E}_{U,L} = \{p(\mathbf{s}), U_{L,\mathbf{s}}\}, \qquad \mathcal{E}'_{U,L} = \{p(\mathbf{s}), \sigma_{\mathbf{s}}U_{L,\mathbf{s}}\}. \tag{18}$$

This ambiguity results in logical Pauli errors, which can induce a logical 1-design for one decoder but not for another. Therefore, we study the formation of logical unitary ($k \geq 2$)-designs separately from 1-designs. While the former come from the intrinsic randomness of the projected ensemble, the latter are affected by the choice of decoder.

To tease out the effect of the decoder choice from the intrinsic randomness in the projected ensemble and connect it to error correction, we first consider two extreme decoders—the optimal decoder and random decoder. We then consider a practical decoder that is intermediate between the two: the minimum weight perfect matching (MWPM) decoder that is a standard choice for the surface code [53,54]. In our application of the MWPM decoder we use no prior information about the noise model. The optimal decoder and random decoder are defined as follows.

(a) The optimal decoder [49] is defined as

$$U_{\text{opt},\mathbf{s}} = \sigma_{\text{opt},\mathbf{s}} U_{L,\mathbf{s}} \tag{19}$$

with

$$\sigma_{\text{opt},\mathbf{s}} = \underset{\sigma\in\{I,X,Y,Z\}}{\arg\max} \, |\text{Tr}(\sigma U_{L,\mathbf{s}})|^2, \tag{20}$$

where $U_{L,\mathbf{s}}$ is the logical unitary obtained from any other decoder, for example MWPM. This choice is relevant to questions about error correction, but not about unitary designs. Indeed, the projected ensemble constructed in this way cannot cover all of $U(2)$ for a trivial reason: the optimal decoder always returns unitaries that are closer to I than to X, Y, or Z; such unitaries make up exactly one quarter of $U(2)$ [see the sketch in Fig. 1(b)]. This prevents the formation of 1-designs, and thus k-designs for any k.

(b) A random decoder is defined as

$$U_{\text{rand},\mathbf{s}} = \sigma_{\text{rand},\mathbf{s}} U_{L,\mathbf{s}}, \tag{21}$$

where $U_{L,\mathbf{s}}$ is the unitary obtained from any other decoder (for example, MWPM or optimal) and $\sigma_{\text{rand},\mathbf{s}}$ is a logical Pauli unitary chosen uniformly from $\{I,X,Y,Z\}$, independently and identically for each syndrome $\mathbf{s}$. This choice is relevant to questions about unitary designs. It is also natural for applications since logical Pauli unitaries can always be implemented transversally. With this "Pauli scrambling" operation, the projected ensemble forms a 1-design with high probability, while formation of k-designs with $k \geq 2$ relies on the intrinsic randomness of syndrome outcomes.

A. Unitary design phase transition

We now investigate convergence of the projected ensemble to a unitary k-design numerically for $k = 2, 3$, using the design distance measure $\Delta^{(k)}$ (see Sec. II A). Specifically, we consider odd-distance-rotated surface codes (see Fig. 2) with the simplest choice of U that meets the requirements of Theorem 1: $U = \prod_j u_j$ with each u_j a single-qubit gate acting on qubit j. To tune the strength of U across the error-correction threshold, we sample each u_j from the Haar measure on $U(2)$ with probability p, and we set $u_j = \mathbb{I}$ otherwise. We denote the resulting measure over the many-body unitary group $U(2^n)$ by μ_p. The parameter $p \in [0, 1]$ is what tunes the strength of U.

The polynomial-time classical algorithm in Ref. [31], based on a free Majorana fermion representation, relies on having only Z rotations on the physical qubits, and fails for generic u_j rotations. Here we propose a tensor-network-based algorithm that can substantially extend the accessible system sizes compared to other existing methods, allowing us to study the unitary design phase transition numerically. We elaborate on the algorithm and investigate its computational complexity in Sec. V.

In Figs. 3(a)–3(d), we show a finite-size scaling analysis of the unitary k-design distance

$$\Delta^{(k)} = \mathop{\mathbb{E}}_{U\sim\mu_p} \left\| \sum_{\mathbf{s}} p(\mathbf{s}) \mathcal{U}_{\text{dec},\mathbf{s}}^{\otimes k} - \Phi_{\text{Haar}}^{(k)} \right\|_{\diamond} \quad \text{for } k = 2, 3, 4, \tag{22}$$

$$\mathcal{U}_{\text{dec},\mathbf{s}}(\cdot) = U_{\text{dec},\mathbf{s}}[\cdot]U_{\text{dec},\mathbf{s}}^{\dagger}, \tag{23}$$

where "dec" stands for the choice of decoder; here we use the random decoder. Data are shown as a function of p (the probability of having a nontrivial physical unitary on each qubit), with varying code distance d from 7 to 15 (always odd). We compute the diamond norm by using the Qiskit package [55], where the diamond norm is formulated as a semidefinite program based on Ref. [56]. The existence of finite-size crossing points for $\Delta^{(k)}$ indicates a unitary design phase transition in the projected ensemble, controlled by p: for $p > p_c$, $\Delta^{(k)}$ decreases in the code distance d, while for $p < p_c$, $\Delta^{(k)}$ increases in the code distance d. The former behavior denotes convergence to a unitary k-design in the thermodynamic limit $d \to \infty$. (See Appendix B for additional data on the distribution of logical unitaries in the two phases, which supports convergence toward the Haar distribution in the design phase.) We refer to the $p > p_c$ regime as a "unitary design phase."

We use the scaling ansatz

$$\Delta^{(k)}(p, d) = F_k[(p - p_c)d^{1/\nu}] \tag{24}$$

to determine the location of the critical point p_c and the correlation length critical exponent ν. From the data collapse in Figs. 3(b), 3(d), 3(f), we locate $p_c \approx 0.87$ and $\nu \approx 1.6$ for all $k = 2$, $k = 3$, and $k = 4$, consistent with a single unitary design phase transition for different moments k. We conjecture that the same critical point should control the convergence of all moments k, thus giving the Haar distribution for $d \to \infty$ when $p > p_c$. We note that the cardinality of the projected ensemble (2^{n-1}, the number of possible syndrome outcomes) allows the formation of unitary designs up to moments $k \leq 2^{n/2}$ [57].

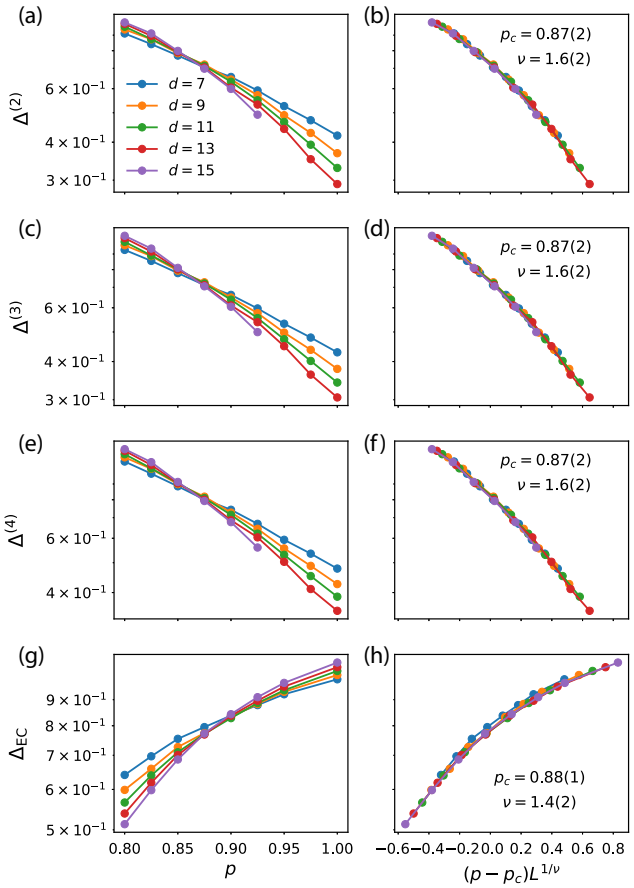


FIG. 3. (a),(c),(e) Unitary k-design distance $\Delta^{(k)}$ for $k = 2, 3, 4$ with the random decoder and (g) logical error rate Δ_{EC} with the optimal decoder as a function of the probability of applying random single-qubit unitaries p. (b),(d),(f),(h) Corresponding scaling collapse of the data. The data are averaged over 384–1280 realizations of random single-qubit unitaries.

B. Error-correction threshold

Using the same simulation method, we can also investigate the optimal error-correction threshold via the logical error rate defined as

$$\Delta_{\mathrm{EC}} = \mathop{\mathbb{E}}_{U\sim\mu_p}\left[\sum_{\mathbf{s}} p(\mathbf{s})\|\mathcal{U}_{\mathrm{opt},\mathbf{s}} - \mathcal{I}\|_{\diamond}\right], \tag{25}$$

where $\mathcal{I}$ is the identity channel, and μ_p is the measure over the unitary group $U(2^n)$ induced by our prescription for sampling the physical gates $\{u_j\}$ [u_j Haar random on $U(2)$ with probability p, and $u_j = \mathbb{I}$ otherwise]. Note the difference with the design distance measures $\Delta^{(k)}$ [Eq. (23)]: here the diamond norm is inside the syndrome average, so Δ_{EC} is small if and only if each logical gate $U_{L,\mathbf{s}}$ is close to the identity.

The numerical results are presented in Figs. 3(e) and 3(f). The data suggest that the logical error rate Δ_{EC} decreases in the code distance d when $p < p_{c,\mathrm{EC}}$, where $p_{c,\mathrm{EC}}$ can be viewed as the coherent error threshold. Similar to the unitary design, we locate the critical point $p_{c,\mathrm{EC}}$ and correlation length exponent ν by finding a data collapse on the scaling ansatz

$$\Delta_{\mathrm{EC}}(p, d) = F[(p - p_{c,\mathrm{EC}})d^{1/\nu}]. \tag{26}$$

The result is $p_{c,\mathrm{EC}} = 0.88(1)$ and $\nu = 1.4(2)$, both of which fall in the tolerance interval of the results for the unitary design phase transition. This provides numerical evidence of a unitary design phase transition for $k \geq 2$ that coincides with the optimal error-correction threshold.

As we argued above, the optimal error-correction threshold poses a lower bound for the formation of unitary designs. Assuming that a unitary design phase exists, one could *a priori* have two distinct thresholds $p_{c,\mathrm{EC}} < p_c$ with an intervening "non-EC, nondesign" phase. Our results indicate that this is not the case. In particular, they suggest that the projected ensemble can flow toward two distinct fixed points as $d \to \infty$: the random Pauli ensemble [58] $\mathcal{E}_{\mathrm{Pauli}} = \{1/4, \mathbb{I}; 1/4, X; 1/4, Y; 1/4, Z\}$ when $p < p_c$, and the Haar-random ensemble $\mathcal{E}_{\mathrm{Haar}}$ when $p > p_c$.

C. Phase transitions under the MWPM decoder

Finally, we study the phase transitions in error correction and the formation of the unitary 1-design under a specific decoder (neither random nor optimal), such as MWPM. For the MWPM decoder, one incurs logical Pauli errors due to the error degeneracy [59], i.e., the ambiguity in the choice of matching. In this case, since the extra logical Pauli errors can form a logical 1-design (but not a 2-design or higher), the error-correction threshold and 1-design transition will be separated from the $(k \geq 2)$-design transition, with $p_{c,\mathrm{MWPM}} \approx 0.20$, much smaller than the $(k \geq 2)$-design threshold $p_c \approx 0.87$, as shown in Fig. 4. The wide discrepancy in thresholds between the MWPM decoder and the optimal decoder arises due to the different information available to the two decoders: our MWPM decoder uses only syndrome data and no prior information on the noise, while the optimal decoder uses full knowledge of the configuration of coherent errors.

V. STAIRCASE CIRCUIT MAPPING AND COMPUTATIONAL PHASE TRANSITION

In this section we present a classical algorithm to decode the surface code under general single-qubit coherent errors, which we used to simulate the projected ensemble of logical unitaries in Sec. IV. Notably, our algorithm undergoes a complexity phase transition at the same value of p_c identified above as the optimal error-correction and unitary design thresholds, with efficient simulation below the threshold and inefficient simulation above the threshold. This complexity transition is due to a measurement-induced entanglement transition in

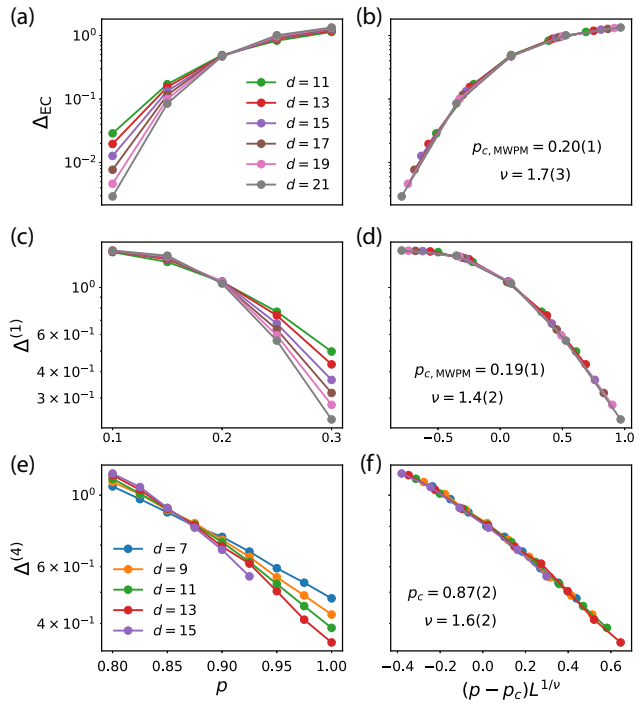


FIG. 4. (a),(b) Logical error rate Δ_{EC}, (c),(d) 1-design distance $\Delta^{(1)}$, and (e),(f) 4-design distance $\Delta^{(4)}$ under the MWPM decoder as a function of the probability of applying random single-qubit unitaries p. Scaling collapse of the data. The data are averaged over 384–1280 realizations of random single-qubit unitaries.

an effective $(1+1)$-dimensional monitored dynamics that appears in the algorithm (see Sec. II C). We stress that our algorithm is of independent interest and could be applied more broadly. It may also be possible to generalize it to the case of incoherent errors [60], which is an interesting direction to pursue in future work.

In Refs. [31,61] an efficient simulation algorithm based on a mapping of the surface code with coherent Z errors to a system of free Majorana fermions was introduced. This is key to the ability to efficiently decode the problem and thus successfully correct errors below the threshold. In Sec. IV we generalized the construction of unitary projected ensembles beyond Z errors, allowing for arbitrary on-site unitaries as well as a wide class of many-body unitaries. In such cases the mapping to free Majorana fermions breaks down, and a new approach is needed.

Our algorithm is based on the "staircase" circuit representation of the encoder (decoder) of surface codes. Surface code states possess long-range (topological) entanglement, and so cannot be prepared by shallow local unitary circuits. Staircase circuits sequentially apply gates across the system, building up the long-range entanglement step by step, and achieve the optimal linear circuit depth lower bound [62,63]. While in practice measurement-based circuits may have an advantage for

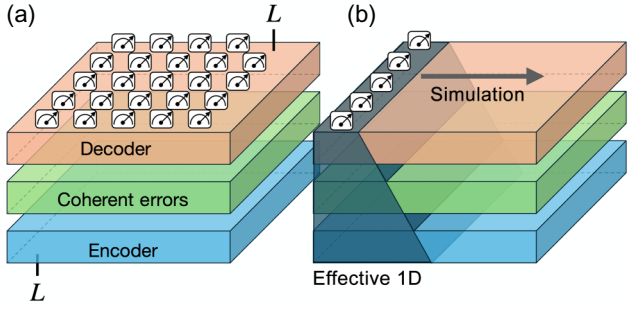


FIG. 5. (a) Schematic of the encoding-decoding circuits with coherent errors. (b) The staircase representation enables the mapping of 2D circuits to effective 1D circuits along the spatial direction, similar to the space-evolving block decimation [46,48].

preparing surface code states [64,65], at the level of classical simulation it will prove useful to focus on staircase encoder circuits. Similarly, staircase decoder circuits can be used to turn the two- and four-qubit measurements of surface code stabilizers into single-qubit measurements. The sequential nature of staircase encoder-decoder circuits, combined with the repeated measurement of individual qubits, allows us to effectively trade a space dimension for a time dimension. This turns the 2D problem with final measurements into a $(1+1)$D dynamics with midcircuit measurements; see Fig. 5. This general strategy was briefly reviewed in Sec. II C and is the basis of the SEBD simulation algorithm [46,48].

A. Warmup: repetition code

Before diving into the algorithm for surface codes, we first consider the repetition code as a toy example. The encoder circuit E of the repetition code is a sequence of controlled-NOT (CNOT) gates arranged in a staircase pattern. The first input qubit contains the state to be encoded, $|\psi\rangle$, while the other $n-1$ qubits store a syndrome string $\mathbf{s} \in \{0,1\}^{n-1}$; E acts on these input states as

$$E[(\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |\mathbf{s}\rangle] = \alpha\,|0,\mathbf{s}\rangle + \beta\,|1,\bar{\mathbf{s}}\rangle\,, \tag{27}$$

where $\bar{\mathbf{s}}$ is the bitwise negation of bitstring $\mathbf{s}$. In words, the output state is a "Schrödinger's cat" of two bitstrings with bond checks $g_i = Z_i Z_{i+1} = (-1)^{s_i}$. The trivial syndrome $\mathbf{s} = \mathbf{0}$ yields repetition code logical states, $\alpha\,|0\rangle^{\otimes n} + \beta\,|1\rangle^{\otimes 1} = |\psi_L\rangle$. It follows that, when we eventually want to measure the parity checks $g_i = Z_i Z_{i+1}$ after acting with a physical unitary U, we may instead apply the decoder circuit $E^\dagger$ followed by computational basis measurements on qubits $i = 2,\ldots,n$, which after application of $E^\dagger$ store the syndrome bits. In all, the protocol is given by $(\mathbb{I} \otimes |\mathbf{s}\rangle\langle\mathbf{s}|)E^\dagger U E(|\psi\rangle \otimes |0\rangle^{\otimes n-1})$, where $\mathbb{I} \otimes |\mathbf{s}\rangle\langle\mathbf{s}|$ projects the syndrome qubits $i = 2,\ldots,n$ onto the syndrome outcome and acts trivially on the logical qubit $i = 1$. This describes

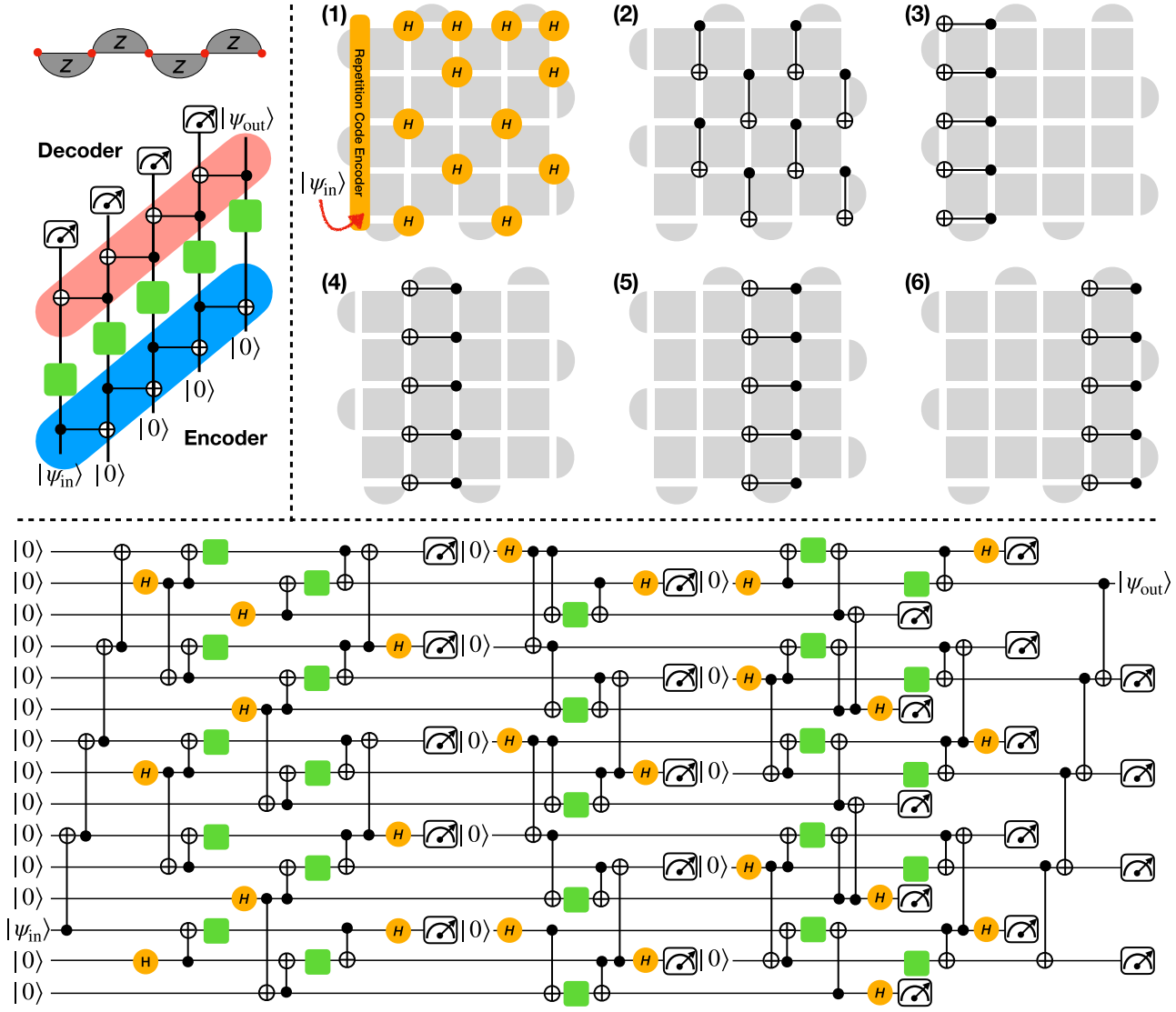


FIG. 6. Upper left panel: repetition code with $d = 5$ and its staircase encoding (blue)-decoding (red) circuit with on-site coherent errors (green). Upper right panel: steps of the staircase encoder for a $d = 5$ surface code. Step (1) includes a depth-d staircase circuit (repetition code encoder), while all the other steps have depth 1. Steps (3)–(6) constitute a staircase of CNOT gates along the horizontal direction, which has total depth $d - 1$. Lower panel: effective $(1 + 1)$D monitored circuit on 3×5 qubits for the encoding-decoding circuit of the $d = 5$ surface code.

a left-to-right sweep of CNOT gates (E), followed by single-qubit gates (U) and then a right-to-left sweep of CNOT gates ($E^\dagger$) and measurements.

It is advantageous to write the whole protocol as a staircase circuit. We can do so by choosing a different decoder circuit, $D = \tilde{E}^\dagger$, where $\tilde{E}$ is simply the mirror reflection of E. This way both E and D proceed left to right. The whole protocol $(|\mathbf{s}\rangle\langle\mathbf{s}| \otimes I)DUE(|\psi\rangle \otimes |\mathbf{0}\rangle)$ is shown diagrammatically in the upper left panel of Fig. 6. It takes in a logical input state in the leftmost qubit, $i = 1$, and returns a logical output state on the rightmost qubit, $i = n$. The tensor network connecting input to output may be seen as the iteration of a unit cell made of only three qubits. At the end of each unit cell, a qubit is measured out and a new auxiliary qubit in the $|0\rangle$ state is introduced, keeping the number of qubits constant and finite. This way we have effectively turned the simulation of a 1D repetition code into a $(0 + 1)$D monitored evolution.

B. Surface code

The surface code, as a hypergraph product of two repetition codes [66], also features a staircase structure. In the upper right panel of Fig. 6, we explicitly show a circuit to encode a $d = 5$ surface code with a staircase of CNOT gates along the horizontal direction. The first step is to encode the input qubit (bottom part of the upper left panel in Fig. 6) into a 1D repetition code along an edge,

with a 1D staircase circuit that sweeps the vertical direction, bottom to top in this case. Then a 2D staircase circuit sweeping the system left to right turns this 1D repetition code logical state into a 2D surface code logical state. The circuit as a whole has linear depth. In analogy to the repetition code, we can choose a decoder circuit $D = \tilde{E}^\dagger$, where $\tilde{E}$ is obtained from E by a spatial inversion, so that the total encoding-decoding circuit takes a staircase form.

Having recast the whole protocol as a staircase circuit followed by one-qubit measurements, we can now view it as a $(1+1)$D monitored evolution, in analogy with the repetition code, but with an additional transverse spatial dimension. This is a version of a "holographic" or "space-time dual" mapping, which leverages the equivalence between space and time in the presence of measurements [44,67–70] to trade a space dimension for a time dimension [see Fig. 5(b)]. The idea is similar to the space-evolving block decimation method for sampling 2D shallow circuits [46,48]; see Sec. II C.

As a result of this mapping, we have an effective 1D description of the problem, which paves the way for matrix product state (MPS) algorithms. The relevant MPS has $N = 3d$ qubits [corresponding to a $3 \times d$ strip of qubits, with the factor of 3 coming from the geometry of the past lightcone in Fig. 5(b)]. See the lower panel of Fig. 6 for an explicit example of a $d = 5$ surface code. Notably, the final (syndrome) measurements are converted to mid-circuit measurements and resets in the dynamics of the MPS. This is especially interesting since midcircuit measurements can lower the state's entanglement entropy and even drive a transition to an area-law-entangled phase (see Sec. II C). The accuracy of the MPS representation relies on the entanglement area law [71].

Similar to the random circuit sampling problem studied in Refs. [46,48], the unitary gates comprising encoder E, "coherent errors" U, and decoder D are competing with the syndrome measurements; this raises the possibility of an entanglement phase transition. Since the number of measurements is fixed by the circuit's geometry, the competition is driven by the strength of coherent errors—the same parameter that, as we saw, drives the unitary design and error-correction transition. It is natural to conjecture an entanglement phase transition in this problem, and thus a complexity transition for the associated classical algorithm, at the same threshold p_c of coherent errors. In the following we investigate this conjecture both numerically and analytically.

C. Entanglement phase transition

In this section we numerically study the entanglement phase transition in the effective 1D system that is used in the above algorithm (Fig. 6, lower panel). The model we consider is the same as in Sec. IV A, where coherent errors are on-site Haar-random unitaries with density p. Because of the structured geometry of this circuit, the tripartite mutual information (the clearest and most direct diagnostic of the entanglement phase transition in one dimension [72]) suffers from large finite-size effects. Therefore, we resort to the coding perspective on the measurement-induced entanglement phase transition [38,39]: we ask whether encoded information can be destroyed by measurements. Concretely, one can consider the behavior of a reference qubit R initially entangled with the monitored system [39]. Let $S_R(t)$ be the entanglement entropy of the reference qubit as a function of time t (in our case in the depth of the monitored circuit). At late time one generally has $S_R(t) \sim \exp(-t/\tau)$, with a purification time τ. The behavior of τ changes sharply at the transition: from $\tau \sim \exp(n)$ in the volume-law phase to $\tau \sim O(1)$ in the area-law phase. At the critical point, τ diverges algebraically as $\tau \sim L^z$ with L the linear size of the system and z a dynamical critical exponent, which is usually 1 [72]. Therefore, in monitored random circuits, one can use τ/L as a diagnostic for the entanglement phase transition.

However, the general picture we just reviewed becomes more subtle in our case. This is due to the fact that our monitored circuits have a special structure, being formed by encoder and decoder circuits of the surface code: the initial state naturally factors into a logical part and a syndrome part. If the reference qubit is entangled with the code's logical qubit, its entanglement with the system will be persistently maximal (one bit), since, based on Theorem 1, the syndrome measurements result in logical unitaries and thus never read out the logical information.

To circumvent this issue, we instead consider entangling the reference qubit to logical degrees of freedom, then running the monitored evolution over time, and eventually measuring out the logical qubit while leaving some syndrome degrees of freedom unmeasured. This setup can be naturally constructed in the MPS representation of the encoding-decoding circuit. We encode a $(d_x + 1) \times d_y$ surface code [73], then treat the first $d_x \times d_y$ qubits (leaving out the last column) as if they formed a surface code patch. We apply coherent errors on this $d_x \times d_y$ surface code, decode it, and measure out the corresponding logical qubit [see Fig. 8(a)]. The remaining column of d_y qubits retains some syndrome information from the larger $[(d_x + 1) \times d_y]$ surface code, and thus can remain entangled with the reference qubit [see Fig. 5(b)]. This circuit is closely related to the first d_x time steps in the $(1+1)$D monitored circuit. It differs only by the application of a repetition code decoder and a logical measurement; these are artificial extra operations that do not alter the phase transition (see Appendix C), but have the advantage of offering an analytical understanding of the connection between the unitary design phase transition and entanglement phase transition (see Sec. V D).

This construction allows us to study the purification of the reference qubit as we sweep across the staircase along

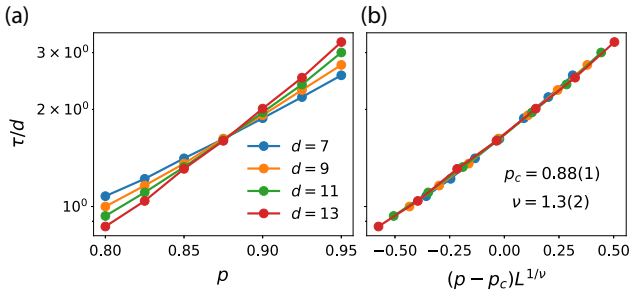


FIG. 7. (a) Purification time scale τ as a function of the probability p of random single-qubit physical unitaries. (b) Scaling collapse of the data. The data are averaged over 100–1000 realizations of both the random coherent errors and syndrome sampling.

one direction—here we choose the horizontal (x) direction. Setting $d_y = d$ and viewing d_x as a variable time scale, the averaged entropy of the reference qubit takes the form $S_R(d_x) \sim \exp(-d_x/\tau)$.

In Fig. 7, we show a finite-size scaling analysis of τ/d for the model considered in Sec. IV, obtained from the MPS simulation of the spacewise dynamics up to $d_x = 3d$ (only $d \le d_x \le 3d$ is used in the fit to avoid early time transient effects). The existence of a finite-size crossing point of τ/d in Fig. 7(a) indicates that $z = 1$ ($\tau \propto d^1$ at criticality). Therefore, we can locate the critical point p_c and correlation length exponent length ν by using a similar scaling ansatz to that in Eq. (24). From the data collapse in Fig. 7(b), we have $p_c = 0.88(1)$ and $\nu = 1.3(2)$, which coincide with the critical point of the unitary design phase transition within tolerance. This indicates a deeper connection between the unitary design phase transition and the entanglement phase transition.

In Appendix C we numerically verify that the purification transition of S_R coincides with the entanglement transition in the 1D MPS in Clifford circuits (replacing Haar-random single-qubit coherent errors by random Clifford coherent errors). There, by using stabilizer simulation, we can compute the tripartite mutual information I_3 in large codes ($d > 100$) and verify that the purification of a reference qubit does signal the entanglement phase transition.

D. Connection between the unitary design phase transition and entanglement phase transition

So far, we have made a numerical observation that the entanglement phase transition in our $(1 + 1)$D monitored dynamics associated with the surface code simulation has the same critical point p_c as the optimal error-correction and unitary design thresholds uncovered in Sec. IV. We now derive an analytical understanding of this connection.

The key technical observation is a mapping of the many-body circuit of Fig. 8(a) onto a three-qubit one, where the

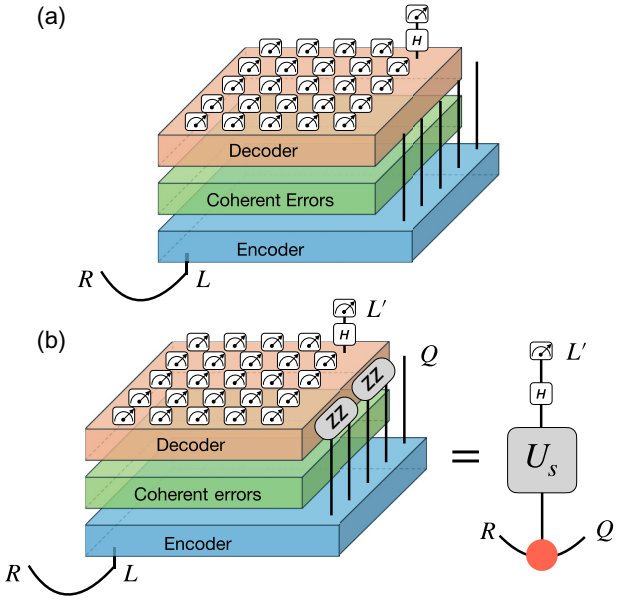


FIG. 8. (a) Schematic of the circuit used in Sec. V C to study dynamical purification of a reference qubit. The reference qubit R is entangled with the logical qubit L of a $(d_x + 1) \times d_y$ surface code, then coherent errors and syndrome measurements are applied only to a $d_x \times d_y$ surface code patch, leaving out a column of qubits on the right. We also measure the logical qubit L' of the $d_x \times d_y$ surface code. (b) Mapping of the circuit to a three-qubit GHZ state. Pairwise ZZ measurements on the qubits in the rightmost column leave out a single qubit Q. Qubits R, L', and Q form a GHZ state. The coherent errors and syndrome measurements result in a unitary $U_{L,\mathbf{s}}$ on L' prior to its measurement.

first qubit represents the reference R, the second is the logical qubit L' (which is being measured in our setup), and the third, Q, represents the unmeasured syndrome degrees of freedom. The three qubits are in an X-basis Greenberger-Horne-Zeilinger (GHZ) state, $|+\rangle^{\otimes 3} + |-\rangle^{\otimes 3}$ [see Fig. 8(b)]; the L' qubit is being acted upon by the logical gate $U_{L,\mathbf{s}}$ and then measured in the X basis. Clearly, if the gate is trivial, $U_{L,\mathbf{s}} \approx \mathbb{I}$, then the GHZ state disentangles into $|\pm\rangle_R |\pm\rangle_Q$ after the measurement, and thus the reference purifies. On the other hand, if the gate is far from the identity, it effectively rotates the measurement away from the X basis and allows for some entanglement between R and Q to survive. This draws a precise connection between the projected ensemble $\mathcal{E}_{U,L} = \{U_{L,\mathbf{s}}\}$ and the entanglement of the reference qubit, and draws an analytical connection between the two transitions.

Next, we explain the mapping in detail. We return to the setup discussed in Sec. V C, where we have a reference qubit entangled with the syndrome degrees of freedom of a $(d_x + 1) \times d_y$ surface code, and we aim to separate that into a $d_x \times d_y$ surface code patch and a $1 \times d_y$ repetition code. It can be shown (see Appendix D) that, by performing two-body ZZ measurements on all but one of the qubits

in the last column [see Fig. 8(b); recall that d_y is odd], one indeed obtains a $d_x \times d_y$ surface code and a d_y repetition code. Calling the repetition code logical qubit Q and the $d_x \times d_y$ surface code logical qubit L', one can show that reference R forms a GHZ state with Q and L' at this stage: $(|+\rangle_R |+\rangle_{L'} |+\rangle_Q + |-\rangle_R |-\rangle_{L'} |-\rangle_Q)/\sqrt{2}$. At this point we apply coherent errors to the $d_x \times d_y$ surface code, followed by decoding and measurement of the syndrome qubits; this acts on logical qubit L' as a unitary drawn from the appropriate projected ensemble $\mathcal{E}_{L,U}$. Finally, we measure out L' in the X basis. This concludes the mapping illustrated in Fig. 8(b).

We note that, although the reference qubit's entanglement in Fig. 8(a) will generally be altered by the ZZ measurements in Fig. 8(b), the qualitative behavior across the phase transition remains the same.

From Sec. IV, we know that the projected ensemble of logical unitaries flows to the random Pauli ensemble $\mathcal{E}_{\text{Pauli}}$ (or the identity ensemble $\mathcal{E}_{\mathbb{I}}$ depending on the decoder) in the thermodynamic limit $d \to \infty$ for $p < p_c$. In this case the logical qubit L' is being measured in the X basis, regardless of the possible applied Pauli unitary; this measurement disentangles the X-basis GHZ state and thus purifies the reference qubit. When $p > p_c$, the projected ensemble flows instead to $\mathcal{E}_{\text{Haar}}$, and application of a random unitary on L' effectively rotates the measurement away from the X basis, leaving some entanglement between R and Q. This can be quantified by the average purity of the reference qubit, i.e.,

$$\langle \mathcal{P} \rangle_R = \tfrac{2}{3}, \tag{28}$$

as can be obtained by Weingarten calculus.

This argument can be adapted to the setting of Fig. 8(a), where we do not apply the extra ZZ measurements on the final column of qubits. The GHZ state on $RL'Q$ can be replaced by a superposition over different ZZ measurement outcomes,

$$|\psi_{\mathbf{s},l}\rangle_{RQM} = \sum_{\mathbf{m}} \sqrt{2p_{\mathbf{m}|\mathbf{s}}} \langle l_{L'}| U_{\mathbf{m},\mathbf{s}} |\text{GHZ}\rangle_{RL'Q} |\mathbf{m}\rangle_M, \tag{29}$$

with l, $\mathbf{s}$, and $\mathbf{m}$ the measurement outcomes of L', the syndrome, and ZZ operators, respectively ($\mathbf{m} \in \{0,1\}^{(d_y-1)/2}$), where we introduced a label M for the latter subsystem and used the fact that $p_l = \frac{1}{2}$. In this case, when $p < p_c$, the reference qubit is purified on average in the thermodynamic limit for the same reason. For $p > p_c$, we show in Appendix E that the averaged purity in the reference qubit satisfies

$$\langle \mathcal{P} \rangle_R \leq \tfrac{2}{3}. \tag{30}$$

Intuitively, the ZZ measurements on average tend to purify R, so if R remains entangled to Q then it was likely to be entangled to QM before M was measured.

Under this perspective, whether the reference qubit is entangled with the rest of the system can be directly related to whether the projected ensemble flows to the identity (Pauli) or Haar distribution. Note that the entanglement (purification) phase transition is independent of the decoder, and thus generally coincides with the $k \geq 2$-design transition.

VI. DISCUSSION

A. Summary

We have introduced a protocol to apply random unitary gates directly on encoded qubits by leveraging the intrinsic randomness of measurement in quantum mechanics. Our protocol is based on the application of unitary transformations on the physical qubits followed by syndrome measurements and correction. Each of these steps can be implemented efficiently on quantum processors, and requires no auxiliary qubits (unlike, e.g., magic state distillation).

We have provided criteria on the quantum error-correcting code and on the physical unitary transformation (Theorem 1) that ensure unitarity of the random logical gates, significantly generalizing previously known results and allowing the realization of arbitrary logical gates. We have then investigated, both numerically and analytically, the emergence of unitary designs in this protocol, uncovering a sharp threshold in the strength p of the applied physical unitaries. When $p > p_c$, we have a "unitary design phase" where, upon taking the code distance $d \to \infty$, the distribution of logical unitaries approaches the Haar-random distribution; when $p < p_c$, all gates in the distribution converge toward the identity, which can be seen as a successful correction of coherent errors, or to random Pauli unitaries, which correspond to an unsuccessful error correction due to a suboptimal decoder. Finally, we have presented a matrix product state algorithm to simulate this protocol, leveraging "staircase" encoder and decoder circuits for the surface code, by which we map the original 2D problem into a $(1+1)$D monitored dynamics. The same threshold p_c that controls the optimal error-correction threshold and unitary design phases also emerges as an entanglement phase transition in this effective monitored dynamics, and thus gives a transition in computational complexity for our MPS algorithm.

B. Outlook

Our results open a number of directions for future investigation and pave the way to several applications. We outline some of them next.

1. Applications of logical unitary designs

Unitary designs have many applications in quantum information science. Our work presents an efficient route

to port these applications to the setting of encoded logical qubits, which is especially interesting as we enter the era of error-corrected quantum processors.

First, we note that some applications such as "Pauli twirling" [74,75] only require a unitary k-design for $k = 1$. On encoded qubits, that can be achieved directly by transversal Pauli unitaries. In these cases the higher level of randomness afforded by our protocol is unnecessary. Furthermore, many protocols [14,16,17,20] require a unitary k-design with $k = 2$ or 3, which is realized by the Clifford group. Some error-correcting codes where our protocol is applicable, such as the odd-distance color code (on the 2D honeycomb lattice), also admit a transversal implementation of the Hadamard and phase gates, which together generate the Clifford group and thus offer a straightforward route to transversal 3-designs.

Our protocol becomes useful for applications requiring ($k = 2$)- or ($k = 3$)-designs on codes that do not admit a transversal implementation of the Clifford group, such as the surface code [76]; or for any applications requiring a 4-design or higher, which entails non-Clifford (or "magic") gates. Below we list some such applications.

Classical shadow tomography is an efficient protocol to learn many properties of quantum states via randomized measurements. One requires a 2-design to get correct results on average, and a 3-design to bound statistical fluctuations [16,17]. Classical shadow tomography is usually formulated for physical qubits, but it is natural to envision applications on fully or partially error-corrected quantum processors where one would need randomized *logical* measurements. On the surface code these could not be implemented transversally, so our method could prove advantageous. A key aspect of the protocol is its division into a "data acquisition" phase, where random unitary gates drawn from a 3-design are applied to the processor before measurement, and a separate "classical postprocessing" phase where the same unitaries are simulated on a classical computer. The feasibility of that simulation is usually guaranteed by the use of Clifford gates. In our case we instead rely on the MPS algorithm presented in Sec. V. This highlights a trade-off between randomness and computational complexity (the MPS algorithm becomes asymptotically inefficient above the threshold, in the unitary design phase).

Randomized benchmarking is a family of practical approaches to estimate the fidelity of unitary gates on a noisy quantum computer [14]. It is based on the application of a long sequence of random gates drawn from a 2-design, followed by a measurement. Comparison of the actual measurement statistics with the ideal one expected under noiseless gates yields an accurate estimate of the fidelity. With our protocol, one could implement randomized benchmarking directly at the logical level. This is interesting in practice since logical gates comprise many physical gates, so learning their fidelity or error models is more challenging. Additionally, *higher-order randomized benchmarking* was proposed in Ref. [77] as a generalization of the aforementioned protocol that makes use of k-designs with $k > 2$ to learn more detailed features of the noise model beyond the fidelity, for example whether or not the noise channel is self-adjoint. The application proposed in Ref. [77] requires exact rather than approximate designs, but we note that in the design phase, error ϵ is suppressed exponentially in distance d, so issues of approximation can likely be avoided by simply scaling up. Finally, we note that, since our protocol implements logical unitaries nondeterministically, it is not directly compatible with standard randomized benchmarking schemes, which rely on deterministic inversion of gates. Instead, our protocol is suitable for inverse-free randomized benchmarking frameworks, such as binary randomized benchmarking [78] and cross-entropy benchmarking [79,80], where inversion gates are replaced by classical postprocessing based on the known gate configuration and measurement outcomes.

Quantum cryptography applications have been a major driver for the introduction of unitary designs [19,81]. Our approach could thus facilitate cryptographic protocols on encoded qubits. Furthermore, since our protocol is not restricted to a specific value of k, it would be interesting to investigate the generation of pseudorandom unitaries [23]. These are unitary ensembles that are computationally indistinguishable from the Haar ensemble even with access to many copies, and are of great interest in quantum cryptography [82].

Random circuit sampling experiments are a key benchmark of computational supremacy in present-day quantum computers [80] and crucially require high levels of randomness that goes beyond the Clifford group (3-design). Even non-Clifford circuits may be classically simulated by so-called Clifford perturbation theory [83] (or Pauli path [84]) methods, if the gates are insufficiently far from Clifford. To remove any loopholes that may be exploited by classical algorithms, it is imperative to build circuits with gates that are as close as possible to genuinely random. Our protocol could find application in next-generation random circuit sampling experiments on partially error-corrected computers. Circuits that combine our highly random single-qubit logical gates with transversal Clifford gates on two logical qubits (e.g., the iSWAP gate, which scrambles very quickly [85]) appear especially promising.

2. Effect of noise

In this work we have focused on the ideal setting where unitary control and measurements are perfect. Where we discussed an error-correction threshold, it was in reference to "coherent errors" (the physical unitaries $\{u_j\}$) that are *applied intentionally* and known by the experimentalist. The performance of this protocol in the presence of

real errors is a natural and crucial question. While a thorough investigation is left to future work, our present results already provide some insight on this.

The "unitary design phase" is defined by the strong randomness in the postmeasurement ensemble of unitaries $\{U_{L,\mathbf{s}}\}$. It is reasonable to expect in that phase that additional (uncontrolled) coherent errors would cause significant damage. The application of unitary designs, as outlined above, requires a classical postprocessing step where data from the noisy experiment are cross-correlated with a simulation of the protocol. If we aim to implement physical unitaries $\{u_j\}$ but, due to, e.g., systematic over- or under-rotation, we implement some other set of gates $\{u'_j\}$, this will lead to a discrepancy between the real and simulated projected ensembles, $\{U^{(\text{expt.})}_{L,\mathbf{s}}\}$ and $\{U^{(\text{sim.})}_{L,\mathbf{s}}\}$. Below the threshold both ensembles should flow to the trivial "identity ensemble" $\mathcal{E}_{\mathbb{I}}$, and thus there is no discrepancy—a signature of the error-correcting phase. But, above the threshold one expects a strong sensitivity to the choice of $\{u_j\}$ and thus a significant difference between experimental and simulated ensembles.

The analysis of incoherent errors is more complex (as it involves a projected ensemble of *channels*), but one can expect similar behavior. Measurement error is the simplest to analyze, as it just leads to a permutation of the labels $\mathbf{s}$ in the projected ensemble; again, we would expect this to cause significant logical errors above the threshold.

Based on these arguments one should not expect the protocol as stated here to be fault tolerant above the threshold. There is thus a trade-off between the amount of randomness in the generated unitary ensemble and the fidelity of the operation (and also the classical computational complexity; see Sec. V). It is interesting to conjecture that the operation at or near the threshold, with multiple repeated rounds of the protocol, might give best results in practice and possibly even fault tolerance. These questions are left to future investigation.

Another important consequence of noise is on the classical simulation of the protocol (necessary for any application of the random logical gates). In the absence of noise, the ensemble of logical gates is independent of how the code is implemented, so in our simulation we are free to use the staircase encoder-decoder circuits that enable our mapping to one dimension. With noise, the specific encoding and syndrome measurement circuits would affect error propagation and would thus need to be included in the simulation, limiting the applicability of our approach. However, to a first approximation, state preparation errors can be treated as channels acting on the input logical state, while measurement errors can be treated as bit flips on the syndrome labels (permutations of the ensemble). In the middle of the protocol, coherent errors can be included straightforwardly. Incoherent errors on the other hand are more challenging, since our MPS method relies on area-law entanglement scaling in pure states. While methods based on matrix product operator representations of the density matrix exist [52,86], there is no guarantee of an area-law scaling of (operator) entanglement in that setting. The possibility of leveraging entanglement transitions under incoherent noise remains an outstanding issue.

3. Extension to general low-density parity check codes and multiple logical qubits

Theorem 1 establishes the unitarity of the projected ensemble for an important set of codes: CSS codes with one logical qubit, odd distance, and whose stabilizers have even Pauli weight. These include the surface code and color code that are practically very relevant to near-term architectures, particularly superconducting qubits where spatial locality is required. It would be interesting to extend our results to more general low-density parity check (LDPC) codes [87]. These are codes with parity checks that are few body, but not necessarily geometrically local. Relaxing geometric locality can be beneficial in terms of scaling of the code rate and distance [88], and is practically relevant to platforms such as trapped ions and neutral atom tweezer arrays.

A challenge lies in the fact that Theorem 1 applies only to codes with one logical qubit ($k = 1$), whereas a key motivation for moving to general LDPC codes is the larger number of logical qubits, potentially even a finite encoding rate ($k \propto n$ as $n \to \infty$). It would be interesting to identify variations of our protocol that can give logical unitaries on codes with $k > 1$. Product constructions of LDPC codes [89] could be a promising route toward this goal, since our understanding of the protocol for repetition and surface codes may be leveraged as a building block to address more complex LDPC codes.

Note added. Recently, we became aware of related works studying the same or related transitions in the surface code. Behrends and Beri [90] mapped the surface code under coherent single-qubit errors to a nonunitary transfer matrix that undergoes an entanglement phase transition. Bao and Anand [91] used a staircase circuit similar to ours and also uncovered an entanglement phase transition. Neither work discusses unitarity of the logical operation or emergence of unitary designs.

ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00112490357. M.J.G. acknowledges support from NSF QLCI Award No. OMA-2120757. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Numerical simulations were performed in part on HPC resources provided by the Texas

Advanced Computing Center (TACC) at the University of Texas at Austin.

DATA AVAILABILITY

The data that support the findings of this article are openly available [92].

APPENDIX A: PROOF OF THEOREM 1

Here we present a proof of Theorem 1. We start by writing the logical density matrix of our $[\![n,1,d]\!]$ stabilizer code in the form

$$\rho_L = \frac{\mathbb{I} + (\mathbf{n}\cdot\boldsymbol{\sigma}_L)}{2}\Pi_{\mathbf{0}}, \tag{A1}$$

where $\boldsymbol{\sigma}_L = (X_L, Y_L, Z_L)$ is the vector of logical operators. We consider the antiunitary time-reversal operator $\mathcal{T} = \mathcal{K}\bigotimes_{j=1}^{n}(iY_j)$, where $\mathcal{K}$ represents complex conjugation. This antiunitary transformation acts by conjugation on Pauli strings P as

$$\mathcal{T}P\mathcal{T}^{-1} = (-1)^{|P|}P, \tag{A2}$$

where $|P|$ is the Pauli weight of the Pauli string (number of nonidentity Pauli matrices in the string).

It is easy to show using conditions (1) and (2) that any representative of any logical operator must have odd weight. It is helpful to prove the following simple fact.

Fact A1. Consider two Pauli operators A and B. If $AB = BA$ then $C = AB$ is a Hermitian Pauli operator of weight $|C| = |A| + |B| \bmod 2$. If $AB = -BA$ then $C = AB/i$ is a Hermitian Pauli operator of weight $|C| = |A| + |B| + 1 \bmod 2$.

Proof. The proof follows straightforwardly from $\mathcal{T}AB\mathcal{T}^{-1} = (-1)^{|A|+|B|}AB = \mathcal{T}C\mathcal{T}^{-1} = (-1)^{|C|}C$ in the first case, and $\mathcal{T}AB\mathcal{T}^{-1} = (-1)^{|A|+|B|}AB = \mathcal{T}iC\mathcal{T}^{-1} = -(-1)^{|C|}iC$ in the second (using the antiunitarity of $\mathcal{T}$). ■

Since stabilizers commute with each logical operation, it follows that, for any stabilizer g and any logical σ_L, we have $|\sigma_L g| = |\sigma_L| + |g| \bmod 2$; using $|g| = 0 \bmod 2$ [condition (1)], this gives $|\sigma_L g| = |\sigma_L| \bmod 2$. So all representatives of a given logical have the same Pauli weight modulo 2. Additionally, since d_x and d_z are both odd, $|X_L|$ and $|Z_L|$ are odd, and since X_L and Z_L anticommute, Fact A1 gives $|Y_L| = |X_L| + |Z_L| + 1 \bmod 2$ and thus $|Y_L|$ is odd as well. We conclude that all nontrivial logical operators have odd Pauli weight.

Based on this fact we can decompose ρ_L as

$$\rho_L = \rho_+ + \rho_-, \tag{A3}$$

with $\rho_+ = \frac{1}{2}\Pi_0$ made exclusively of even-weight Pauli operators and $\rho_- = \frac{1}{2}\Pi_0(\mathbf{n}\cdot\boldsymbol{\sigma}_L)$ made exclusively of odd-weight Pauli operators: $\mathcal{T}\rho_\pm\mathcal{T}^{-1} = \pm\rho_\pm$.

Condition (3) states that $[U,\mathcal{T}] = 0$; therefore, the state after application of U becomes

$$U\rho_L U^\dagger = U\rho_+U^\dagger + U\rho_-U^\dagger = \rho'_+ + \rho'_- \tag{A4}$$

with $\rho'_\pm = U\rho_\pm U^\dagger$ such that $\mathcal{T}\rho'_\pm\mathcal{T}^{-1} = \pm\rho'_\pm$: U does not mix the time-reversal-even and time-reversal-odd sectors. Now the probability of measuring syndrome $\mathbf{s}$ on state $U\rho_L U^\dagger$ is

$$p(\mathbf{s}) = \mathrm{Tr}(\Pi_{\mathbf{s}}U\rho_L U^\dagger) = \mathrm{Tr}(\Pi_{\mathbf{s}}\rho'_+) + \mathrm{Tr}(\Pi_{\mathbf{s}}\rho'_-) \tag{A5}$$

with $\Pi_{\mathbf{s}}$ the projector on syndrome $\mathbf{s}$. It can be easily seen that

$$\Pi_{\mathbf{s}} = \prod_{i=1}^{n-1}\frac{\mathbb{I} + (-1)^{s_i}g_i}{2} \tag{A6}$$

contains only even-weight operators (stabilizers); therefore, $\rho'_-\Pi_{\mathbf{s}}$ contains only odd-weight operators (logicals). As the latter are traceless, we get

$$p(\mathbf{s}) = \mathrm{Tr}(\Pi_{\mathbf{s}}\rho'_+) = \tfrac{1}{2}\mathrm{Tr}(\Pi_{\mathbf{0}}U\Pi_{\mathbf{s}}U^\dagger), \tag{A7}$$

which is independent of the input logical state. In words, all the logical information is stored in odd-weight operators, but the syndrome measurements only probe even-weight operators, and so learn no logical information. At this point Theorem 1 follows from application of Fact 1.

APPENDIX B: ADDITIONAL DATA ON THE DISTRIBUTION OF LOGICAL UNITARIES

In this appendix we provide more fine-grained data on the distributions of logical unitaries in both phases. This gives additional evidence for our claim of unitary designs above the error-correction threshold.

To simplify our analysis, we consider deterministic, uniform coherent errors $U_i = \exp(i\alpha\mathbf{n}\cdot\boldsymbol{\sigma}_i)$, where $\mathbf{n} = (1,1,1)/\sqrt{3}$ and $\boldsymbol{\sigma}_i = (X_i, Y_i, Z_i)$. The angle α plays a similar role to the density of coherent errors p used in the main text and tunes a design (error-correction) phase transition in the projected ensemble. Data shown in Fig. 9 on the 4-design distance $\Delta^{(4)}$ indicate a design phase above $\alpha \gtrsim 0.18\pi$. Note that, for both $\alpha = 0$ and $\alpha = \pi/3$, we have Clifford coherent errors, which cannot form more than 1-designs (see Appendix C), so an upper limit to the

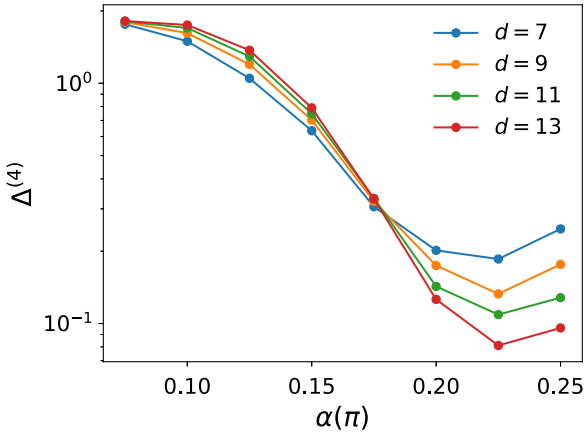


FIG. 9. Unitary 4-design distance $\Delta^{(4)}$ with random decoder for transversal coherent error $U_i = \exp(i\alpha \mathbf{n} \cdot \boldsymbol{\sigma}_i)$ with $\mathbf{n} = (1,1,1)/\sqrt{3}$. Data are sampled over 10 000 instances.

design phase is expected before $\alpha = \pi/3$, and hence the nonmonotonic behavior of $\Delta^{(4)}$ in Fig. 9.

To investigate the unitary distributions in both phases, we choose $\alpha = 0.15\pi$ as representative of the error-correcting phase and $\alpha = 0.225\pi$ as representative of the unitary-design phase. We parametrize the logical unitary in terms of three angles $\beta_{\mathbf{s}}$, $\theta_{\mathbf{s}}$, and $\phi_{\mathbf{s}}$ as

$$U(\mathbf{s}) = \exp(i\beta_{\mathbf{s}} \mathbf{n}_{\mathbf{s}} \cdot \boldsymbol{\sigma}_L), \tag{B1}$$

$$\mathbf{n}_{\mathbf{s}} = (\sin\theta_{\mathbf{s}} \cos\varphi_{\mathbf{s}}, \sin\theta_{\mathbf{s}} \sin\varphi_{\mathbf{s}}, \cos\theta_{\mathbf{s}}). \tag{B2}$$

For the Haar distribution, the probability density function of (β, θ, φ) is given by

$$p(\beta, \theta, \varphi) \propto \sin^2\beta \sin\theta. \tag{B3}$$

In Fig. 10, we compare the marginal distributions of β, θ, and φ between the Haar ensemble and the projected ensemble at $\alpha = 0.15\pi$ and $\alpha = 0.225\pi$. The former clearly shows deviations from the Haar prediction in the form of sharp peaks around multiples of $\pi/2$ for each angle, which are characteristic of Pauli unitaries; the peaks become sharper as distance d increases from 7 to 13, consistent with convergence toward a random Pauli distribution. On the contrary, for $\alpha = 0.225\pi$, we see close

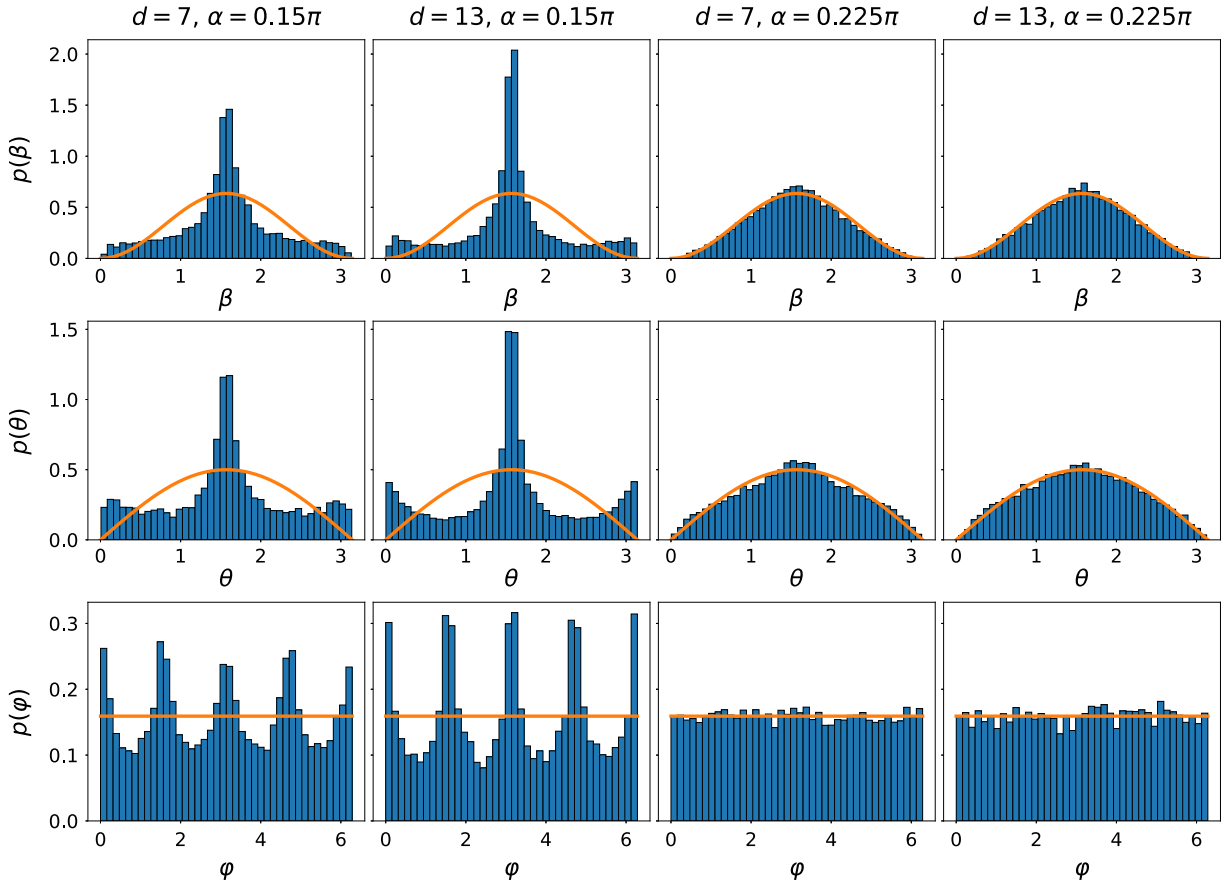


FIG. 10. Distributions of unitary parameters β, θ, ϕ in Eq. (B1) in the projected ensembles for code distances $d = 7$ and 13, and for coherent error strengths $\alpha = 0.15\pi$ (below the threshold) and 0.225π (above the threshold). The probability density functions for the Haar distribution (B3) are shown with orange lines. Data are sampled over 10 000 syndromes.

agreement with the Haar prediction already at $d = 7$, with further improvement (visible especially in θ) as we scale up to $d = 13$. This provides fine-grained evidence of the approach to the Haar distribution in the design phase.

APPENDIX C: CLIFFORD SIMULATION OF THE ERROR-CORRECTION THRESHOLD AND ENTANGLEMENT TRANSITION

In this appendix, we numerically investigate the logical unitary design phase transition and entanglement phase transition in Clifford circuits, where we replace the Haar-random coherent errors by random Clifford ones. Since the encoding and decoding circuits are themselves Clifford, the overall circuit is amenable to large-scale classical simulation via stabilizer formalism [92]. Part of our simulation uses the *QuantumClifford.jl* package [93].

We compute the tripartite mutual information I_3 as a diagnostic of the entanglement phase transition [72], with

$$I_3 = S(A) + S(B) + S(C) - S(A \cup B) - S(A \cup C) - S(B \cup C) + S(A \cup B \cup C), \quad \text{(C1)}$$

where $S(X)$ is the von Neumann entropy of region X and A, B, C are contiguous subsystems of linear size $(d-1)/4$.

An important feature of the projected ensemble of logical unitaries in the Clifford case is that it can form at most a 1-design. This follows from its limited cardinality: at most four distinct elements belong to $\mathcal{E}_L$. One can show this by writing the Choi state associated with a logical unitary $U_{L,\mathbf{s}}$:

$$|U(\mathbf{s})\rangle = \frac{1}{\sqrt{2}} \left(|0_R\rangle U_L(\mathbf{s})|0_L\rangle + |1_R\rangle U_L(\mathbf{s})|1_L\rangle\right). \quad \text{(C2)}$$

This is a stabilizer state of two qubits and, as such, can be represented by two stabilizer generators $g_1 = P_{1,R}P_{1,L}$, $g_2 = P_{2,R}P_{2,L}$ where $P_{1/2,R/L}$ are Pauli matrices on the two qubits that we call L and R. The unitarity of U_L implies that this state is maximally entangled, i.e., $P_{1,R/L} \neq P_{2,R/L}$. For stabilizer states, the outcome of projective measurements can only affect the "phase bits" in the stabilizer tableau, i.e., the ± 1 signs in front of each stabilizer generator. So, for any given syndrome $\mathbf{s}$, we have

$$|U(\mathbf{s})\rangle\langle U(\mathbf{s})| \in \left\{ \frac{\mathbb{I} \pm g_1}{2} \frac{\mathbb{I} \pm g_2}{2} \right\}, \quad \text{(C3)}$$

which, as claimed, contains only four elements. More specifically, we have

$$U(\mathbf{s}) = CP(\mathbf{s}), \quad \text{(C4)}$$

where C is some Clifford gate independent of $\mathbf{s}$ and $P(\mathbf{s})$ is a Pauli gate that may depend on $\mathbf{s}$. As a result, the ensemble consisting of $CP(\mathbf{s})$ can only form a unitary 1-design.

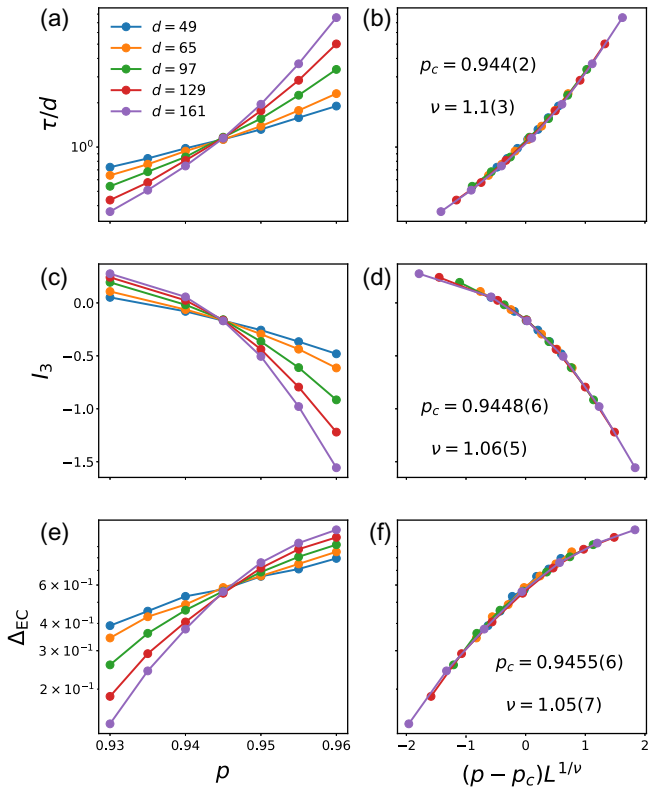


FIG. 11. (a),(b) Purification times τ, (c),(d) tripartite mutual information I_3, (e),(f) logical error rate Δ_{EC} under the optimal decoder as a function of the probability p of applying random single-qubit Clifford operators. Panels (b), (d), and (f) represent the scaling collapse of the data in (a), (c), and (e). The data are averaged over 5120 realizations.

We summarize our numerical results for Clifford simulation in Fig. 11 on purification time τ, tripartite mutual information I_3, and logical error rate Δ_{EC}. Their critical points are all around $p_c \approx 0.945$, indicating the coincidence of entanglement, purification, error-correction, and unitary design phase transitions.

APPENDIX D: DECOMPOSITION OF SURFACE CODES

Here we explain the decomposition of the surface codes used in Sec. V D and Fig. 8(b) to connect unitary design and entanglement phases. The main statement is that, after some ZZ measurements (or XX depending on the orientation of the surface code) on an edge, the stabilizers of a $(d_x + 1) \times d_y$ surface code can be decomposed into stabilizers of a $d_x \times d_y$ surface code and a $1 \times d_y$ repetition code. Different measurement outcomes correspond to the different signs of the stabilizers at the edge.

In Fig. 12, we show this explicitly for the case $d_x = d_y = 5$. At the same time, if we start from a Bell pair state between the logical L of the $(d_x + 1) \times d_y$ surface code and

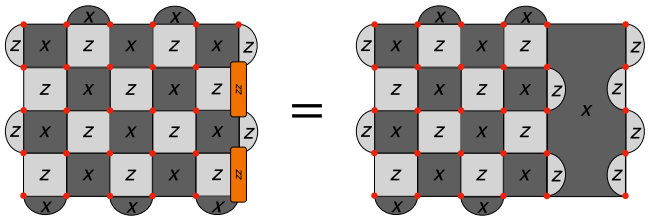


FIG. 12. Decomposition of a 6×5 surface code into a 5×5 surface code and $d = 5$ repetition code after some edge ZZ measurements.

a reference qubit R,

$$\rho_{RL} \propto \frac{\mathbb{I} + X_R X_L}{2} \frac{\mathbb{I} + Z_R Z_L}{2}, \tag{D1}$$

after the decomposition, we have a three-qubit X-basis GHZ state over reference R, the logical L' of the $d_x \times d_y$ surface code, and the logical Q of the $1 \times d_y$ repetition code:

$$\rho_{RL'Q} \propto \frac{\mathbb{I} + X_R X_{L'}}{2} \frac{\mathbb{I} + Z_R Z_{L'} Z_Q}{2} \frac{\mathbb{I} + X_{L'} X_Q}{2}. \tag{D2}$$

Here we have used the fact that $X_L = X_{L'}$, $Z_L = Z_{L'} Z_Q$, and the last $X_{L'} X_Q$ stabilizer comes from the large X plaquette in Fig. 12 (that originates by merging smaller X plaquettes due to the ZZ measurements).

APPENDIX E: UPPER BOUND OF THE AVERAGED PURITY IN THE REFERENCE QUBIT

Here we prove Eq. (30) for the state in Fig. 8(a) in the design phase. The state is explicitly given by Eq. (29):

$$|\psi_{\mathbf{s}l}\rangle_{RQM} = \sum_{\mathbf{m}} \sqrt{2 p_{\mathbf{m}|\mathbf{s}}} \langle l_{L'} | U_{\mathbf{m},\mathbf{s}} | \mathrm{GHZ} \rangle_{RL'Q} |\mathbf{m}\rangle_M \tag{E1}$$

with l, $\mathbf{s}$, $\mathbf{m}$ the measurement outcomes of L', syndrome, and ZZ operators on the last column, respectively, and $p_{\mathbf{m}|\mathbf{s}}$ the conditional probability of $\mathbf{m}$ given $\mathbf{s}$.

The density matrix of the reference qubit R is

$$\begin{aligned} \rho(\mathbf{s}, l)_R &= \mathrm{Tr}_{QM}(|\psi_{\mathbf{s}l}\rangle \langle \psi_{\mathbf{s}l}|) \\ &= \sum_{l'=\pm} |l'\rangle\langle l'|_R \left(\sum_{\mathbf{m}} p_{\mathbf{m}|\mathbf{s}} |\langle l| U_{\mathbf{m},\mathbf{s}} |l'\rangle|^2 \right). \end{aligned} \tag{E2}$$

The averaged purity thus reads

$$\begin{aligned} \langle \mathcal{P} \rangle_R &= \frac{1}{2} \sum_{\mathbf{s},l,l'} p_{\mathbf{s}} \left(\sum_{\mathbf{m}} p_{\mathbf{m}|\mathbf{s}} |\langle l| U_{\mathbf{m},\mathbf{s}} |l'\rangle|^2 \right)^2 \\ &\le \frac{1}{2} \sum_{\mathbf{s},l,l'} p_{\mathbf{s}} \sum_{\mathbf{m}} p_{\mathbf{m}|\mathbf{s}} |\langle l| U_{\mathbf{m},\mathbf{s}} |l'\rangle|^4 \\ &= \frac{1}{2} \sum_{\mathbf{m},l,l'} p_{\mathbf{m}} \sum_{\mathbf{s}} p_{\mathbf{s}|\mathbf{m}} |\langle l| U_{\mathbf{m},\mathbf{s}} |l'\rangle|^4, \end{aligned} \tag{E3}$$

where we have used the fact that $p_l = \frac{1}{2}$ and the Cauchy-Schwarz inequality $(\sum_{\mathbf{m}} a_{\mathbf{m}} b_{\mathbf{m}})^2 \le (\sum_{\mathbf{m}} a_{\mathbf{m}}^2)(\sum_{\mathbf{m}} b_{\mathbf{m}}^2)$ for vectors $a_{\mathbf{m}} = \sqrt{p_{\mathbf{m}|\mathbf{s}}}$ and $b_{\mathbf{m}} = \sqrt{p_{\mathbf{m}|\mathbf{s}}} |\langle l| U_{\mathbf{m},\mathbf{s}} |l'\rangle|^2$. We have also used the definition of the conditional probability rule $p_{\mathbf{m}|\mathbf{s}} p_{\mathbf{s}} = p_{\mathbf{s}|\mathbf{m}} p_{\mathbf{m}}$.

In the design phase, for all $\mathbf{m}$, the ensemble $\mathcal{E}_{\mathbf{m}} \equiv \{p_{\mathbf{s}|\mathbf{m}}, U_{\mathbf{m},\mathbf{s}}\}$ flows to the Haar distribution, since different $\mathbf{m}$ differ only by the sign of certain stabilizers in state $\rho_{RL'Q}$, Eq. (D2), which affects the ensemble by at most an overall Pauli rotation. Replacing $\sum_{\mathbf{s}} p_{\mathbf{s}|\mathbf{m}} |\langle l| U_{\mathbf{m},\mathbf{s}} |l'\rangle|^4 \to \mathbb{E}_{U \sim \mathrm{Haar}} [|\langle l| U |l'\rangle|^4] = \frac{1}{3}$ on the right-hand side of Eq. (E3) yields $\langle \mathcal{P} \rangle_R \le \frac{2}{3}$.

[1] B. Eastin, Restrictions on transversal encoded quantum gate sets, Phys. Rev. Lett. **102**, 110502 (2009).

[2] E. Knill, Fault-tolerant postselected quantum computation: Schemes, ArXiv:quant-ph/0402171.

[3] S. Bravyi, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[4] M. B. Hastings, Distillation with sublogarithmic overhead, Phys. Rev. Lett. **120**, 050504 (2018).

[5] A. Wills, M.-H. Hsieh, and H. Yamasaki, Constant-overhead magic state distillation, ArXiv:2408.07764.

[6] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, H. Pichler, F. G. S. L. Brandão, S. Choi, and M. Endres, Preparing random states and benchmarking with many-body quantum chaos, Nature **613**, 468 (2023).

[7] J. S. Cotler, D. K. Mark, H.-Y. Huang, F. Hernández, J. Choi, A. L. Shaw, M. Endres, and S. Choi, Emergent quantum state designs from individual many-body wave functions, PRX Quantum **4**, 010311 (2023).

[8] W. W. Ho and S. Choi, Exact emergent quantum state designs from quantum chaotic dynamics, Phys. Rev. Lett. **128**, 060601 (2022).

[9] M. Ippoliti and W. W. Ho, Solvable model of deep thermalization with distinct design times, Quantum **6**, 886 (2022).

[10] X. Turkeshi and P. Sierant, Error-resilience phase transitions in encoding-decoding quantum circuits, Phys. Rev. Lett. **132**, 140401 (2024).

[11] P. Niroula, C. D. White, Q. Wang, S. Johri, D. Zhu, C. Monroe, C. Noel, and M. J. Gullans, Phase transition in magic with random quantum circuits, Nat. Phys. **20**, 1786 (2024).

[12] J. Behrends, F. Venn, and B. Béri, Surface codes, quantum circuits, and entanglement phases, Phys. Rev. Res. **6**, 013137 (2024).

[13] F. Venn, J. Behrends, and B. Béri, Coherent-error threshold for surface codes from Majorana delocalization, Phys. Rev. Lett. **131**, 060603 (2023).

[14] E. Knill, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

[15] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[16] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[17] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).

[18] A. A. Mele, Introduction to Haar measure tools in quantum information: A beginner's tutorial, Quantum **8**, 1340 (2024).

[19] A. Ambainis and J. Emerson, in *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)* (IEEE Computer Society, Los Alamitos, CA, USA, 2007), p. 129.

[20] C. Dankert, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[21] D. A. Roberts and B. Yoshida, Chaos and complexity by design, J. High Energy Phys. **2017**, 121 (2017).

[22] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits, ArXiv:1905.12053.

[23] T. Schuster, J. Haferkamp, and H.-Y. Huang, Random unitaries in extremely low depth, Science **389**, 92 (2025).

[24] M. Ippoliti and W. W. Ho, Dynamical purification and the emergence of quantum state designs from the projected ensemble, PRX Quantum **4**, 030322 (2023).

[25] T. Bhore, J.-Y. Desaules, and Z. Papic, Deep thermalization in constrained quantum systems, Phys. Rev. B **108**, 104317 (2023).

[26] D. K. Mark, F. Surace, A. Elben, A. L. Shaw, J. Choi, G. Refael, M. Endres, and S. Choi, Maximum entropy principle in deep thermalization and in Hilbert-space ergodicity, Phys. Rev. X **14**, 041051 (2024).

[27] A. Chan and A. De Luca, Projected state ensemble of a generic model of many-body quantum chaos, J. Phys. A: Math. Theoret. **57**, 405001 (2024).

[28] M. Lucas, L. Piroli, J. De Nardis, and A. De Luca, Generalized deep thermalization for free fermions, Phys. Rev. A **107**, 032215 (2023).

[29] C. Liu, Q. C. Huang, and W. W. Ho, Deep thermalization in Gaussian continuous-variable quantum systems, Phys. Rev. Lett. **133**, 260401 (2024).

[30] R.-A. Chang, H. Shrotriya, W. W. Ho, and M. Ippoliti, Deep thermalization under charge-conserving quantum dynamics, PRX Quantum **6**, 020343 (2025).

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

[32] M. P. A. Fisher, V. Khemani, A. Nahum, and S. Vijay, Random quantum circuits, Annu. Rev. Condens. Matter Phys. **14**, 335 (2023).

[33] A. C. Potter and R. Vasseur, in *Entanglement in Spin Chains: From Theory to Quantum Technology Applications*, Quantum Science and Technology, edited by A. Bayat, S. Bose, and H. Johannesson (Cham, Switzerland, 2022), p. 211.

[34] B. Skinner, J. Ruhman, and A. Nahum, Measurement-induced phase transitions in the dynamics of entanglement, Phys. Rev. X **9**, 031009 (2019).

[35] Y. Li, X. Chen, and M. P. A. Fisher, Quantum Zeno effect and the many-body entanglement transition, Phys. Rev. B **98**, 205136 (2018).

[36] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, Quantum error correction in scrambling dynamics and measurement-induced phase transition, Phys. Rev. Lett. **125**, 030505 (2020).

[37] Y. Bao, S. Choi, and E. Altman, Theory of the phase transition in random unitary circuits with measurements, Phys. Rev. B **101**, 104301 (2020).

[38] M. J. Gullans and D. A. Huse, Dynamical purification phase transition induced by quantum measurements, Phys. Rev. X **10**, 041020 (2020).

[39] M. J. Gullans and D. A. Huse, Scalable probes of measurement-induced criticality, Phys. Rev. Lett. **125**, 070606 (2020).

[40] R. Fan, S. Vijay, A. Vishwanath, and Y.-Z. You, Self-organized error correction in random unitary circuits with measurement, Phys. Rev. B **103**, 174309 (2021).

[41] Y. Li and M. P. A. Fisher, Statistical mechanics of quantum error correcting codes, Phys. Rev. B **103**, 104306 (2021).

[42] Y. Li, Y. Zou, P. Glorioso, E. Altman, and M. P. A. Fisher, Cross entropy benchmark for measurement-induced phase transitions, Phys. Rev. Lett. **130**, 220404 (2023).

[43] C. Noel, P. Niroula, D. Zhu, A. Risinger, L. Egan, D. Biswas, M. Cetina, A. V. Gorshkov, M. J. Gullans, D. A. Huse, and C. Monroe, Measurement-induced quantum phases realized in a trapped-ion quantum computer, Nat. Phys. **18**, 760 (2022).

[44] J. C. Hoke, *et al.*, Measurement-induced entanglement and teleportation on a noisy quantum processor, Nature **622**, 481 (2023).

[45] T. Vovk and H. Pichler, Entanglement-optimal trajectories of many-body quantum Markov processes, Phys. Rev. Lett. **128**, 243601 (2022).

[46] Z. Cheng and M. Ippoliti, Efficient sampling of noisy shallow circuits via monitored unraveling, PRX Quantum **4**, 040326 (2023).

[47] Z. Chen, Y. Bao, and S. Choi, Optimized trajectory unraveling for classical simulation of noisy quantum dynamics, Phys. Rev. Lett. **133**, 230403 (2024).

[48] J. C. Napp, R. L. La Placa, A. M. Dalzell, F. G. S. L. Brandão, and A. W. Harrow, Efficient classical simulation of random shallow 2D quantum circuits, Phys. Rev. X **12**, 021021 (2022).

[49] F. Venn and B. Béri, Error-correction and noise-decoherence thresholds for coherent errors in planar-graph surface codes, Phys. Rev. Res. **2**, 043412 (2020).

[50] Since any two g_i, g_j commute, $g_i g_j$ is a Hermitian Pauli operator; thus, $\mathcal{T} g_i g_j \mathcal{T}^{-1} = (-1)^{|g_i g_j|}$, but also $\mathcal{T} g_i g_j \mathcal{T}^{-1} = \mathcal{T} g_i \mathcal{T}^{-1} \mathcal{T} g_j \mathcal{T}^{-1} = (-1)^{|g_i|+|g_j|} g_i g_j$. Thus, $|g_i g_j| = |g_i| + |g_j| \bmod 2$. So if every element in a generating set $\{g_i\}$ has even Pauli weight then each element of the stabilizer group does.

[51] This subgroup is isomorphic to the compact symplectic group $\mathrm{USp}(2^{n-1})$, of dimension $2^{n-1}(2^n+1)$. This is defined as $\{U \in U(2^n) : U^T \Omega U = \Omega\}$ with Ω a standard symplectic form—a real antisymmetric matrix obeying $\Omega^2 = -I$ and $\Omega\Omega^T = \Omega^T\Omega = I$. One can rewrite $[U, \mathcal{T}] = 0$ as $U^T \Omega U = \Omega$, where $\Omega = \prod_{j=1}^{n} iY_j$ is real antisymmetric and obeys $\Omega^2 = (-1)^n I$ and $\Omega\Omega^T = \Omega^T\Omega = I$. Finally, conditions (1) and (2) imply that the code size n is odd (e.g., the operator $\prod_{j=1}^{n} X_j$ commutes with all stabilizers and anticommutes with Z_L, so it is a nontrivial logical of weight n), giving $\Omega^2 = -I$.

[52] A. S. Darmawan, Optimal adaptation of surface-code decoders to local noise, ArXiv:2403.08706.

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

[54] A. G. Fowler, Towards practical classical processing for the surface code, Phys. Rev. Lett. **108**, 180501 (2012).

[55] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, ArXiv:2405.08810.

[56] J. Watrous, Simpler semidefinite programs for completely bounded norms, ArXiv:1207.5726.

[57] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).

[58] This assumes a random decoder. With a specific decoder like MWPM we have a trivial "identity" ensemble $\mathcal{E}_{\mathbb{I}} = \{1, \mathbb{I}\}$ below a decoder-dependent threshold $p < p_{c,\mathrm{dec}}$, as shown in Fig. 1(b).

[59] T. M. Stace and S. D. Barrett, Error correction and degeneracy in surface codes suffering loss, Phys. Rev. A **81**, 022317 (2010).

[60] J. Behrends and B. Béri, The surface code beyond Pauli channels: Logical noise coherence, information-theoretic measures, and errorfield-double phenomenology, ArXiv:2412.21055.

[61] Y. Suzuki, K. Fujii, and M. Koashi, Efficient simulation of quantum error correction under coherent error based on the nonunitary free-fermionic formalism, Phys. Rev. Lett. **119**, 190503 (2017).

[62] K. J. Satzinger, *et al.*, Realizing topologically ordered states on a quantum processor, Science **374**, 1237 (2021).

[63] O. Higgott, M. Wilson, J. Hefford, J. Dborin, F. Hanif, S. Burton, and D. E. Browne, Optimal local unitary encoding circuits for the surface code, Quantum **5**, 517 (2021).

[64] M. Iqbal, N. Tantivasadakarn, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, A. Hankin, N. Hewitt, C. V. Horst, M. Matheny, T. Mengle, B. Neyenhuis, A. Vishwanath, M. Foss-Feig, R. Verresen, and H. Dreyer, Topological order from measurements and feed-forward on a trapped ion quantum computer, Commun. Phys. **7**, 1 (2024).

[65] M. Foss-Feig, A. Tikku, T.-C. Lu, K. Mayer, M. Iqbal, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, A. Hankin, N. Hewitt, C. V. Horst, M. Matheny, T. Mengle, B. Neyenhuis, H. Dreyer, D. Hayes, T. H. Hsieh, and I. H. Kim, Experimental demonstration of the advantage of adaptive quantum circuits, ArXiv:2302.03029.

[66] J.-P. Tillich and G. Zemor, Quantum LDPC codes with positive rate and minimum distance proportional to the square root of the blocklength, IEEE Trans. Inf. Theory **60**, 1193 (2014).

[67] M. Foss-Feig, D. Hayes, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, J. M. Pino, and A. C. Potter, Holographic quantum algorithms for simulating correlated spin systems, Phys. Rev. Res. **3**, 033002 (2021).

[68] M. Ippoliti, T. Rakovszky, and V. Khemani, Fractal, logarithmic, and volume-law entangled nonthermal steady states via spacetime duality, Phys. Rev. X **12**, 011045 (2022).

[69] T.-C. Lu and T. Grover, Spacetime duality between localization transitions and measurement-induced transitions, PRX Quantum **2**, 040319 (2021).

[70] S. Anand, J. Hauschild, Y. Zhang, A. C. Potter, and M. P. Zaletel, Holographic quantum simulation of entanglement renormalization circuits, PRX Quantum **4**, 030334 (2023).

[71] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Entropy scaling and simulability by matrix product states, Phys. Rev. Lett. **100**, 030504 (2008).

[72] A. Zabalo, M. J. Gullans, J. H. Wilson, S. Gopalakrishnan, D. A. Huse, and J. H. Pixley, Critical properties of the measurement-induced transition in random quantum circuits, Phys. Rev. B **101**, 060301(R) (2020).

[73] Here d_x, d_y refer to the linear dimensions of the surface code.

[74] W. Dür, M. Hein, J. I. Cirac, and H.-J. Briegel, Standard forms of noisy quantum operations via depolarization, Phys. Rev. A **72**, 052326 (2005).

[75] Z. Cai and S. C. Benjamin, Constructing smaller Pauli twirling sets for arbitrary error channels, Sci. Rep. **9**, 11281 (2019).

[76] B. J. Brown, K. Laubscher, M. S. Kesselring, and J. R. Wootton, Poking holes and cutting corners to achieve Clifford gates with the surface code, Phys. Rev. X **7**, 021029 (2017).

[77] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, K. Zuo, S. Tamate, Y. Tabuchi, and Y. Nakamura, Quantum circuits for exact unitary t-designs and applications to higher-order randomized benchmarking, PRX Quantum **2**, 030339 (2021).

[78] J. Hines, D. Hothem, R. Blume-Kohout, B. Whaley, and T. Proctor, Fully scalable randomized benchmarking without motion reversal, PRX Quantum **5**, 030334 (2024).

[79] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, Nat. Phys. **14**, 595 (2018).

[80] F. Arute, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[81] C. Lancien and C. Majenz, Weak approximate unitary designs and applications to quantum encryption, Quantum **4**, 313 (2020).
[82] Z. Ji, Y.-K. Liu, and F. Song, in *Advances in Cryptology—CRYPTO 2018*, edited by H. Shacham and A. Boldyreva (Cham, 2018), p. 126.
[83] T. Begusic and G. K.-L. Chan, Real-time operator evolution in two and three dimensions via sparse Pauli dynamics, PRX Quantum **6**, 020302 (2025).
[84] G. González-García, J. I. Cirac, and R. Trivedi, Pauli path simulations of noisy quantum circuits beyond average case, Quantum **9**, 1730 (2025).
[85] X. Mi, *et al.*, Information scrambling in quantum circuits, Science **374**, 1479 (2021).
[86] A. S. Darmawan and D. Poulin, Tensor-network simulations of the surface code under realistic noise, Phys. Rev. Lett. **119**, 040502 (2017).
[87] N. P. Breuckmann and J. N. Eberhardt, Quantum low-density parity-check codes, PRX Quantum **2**, 040101 (2021).
[88] S. Bravyi and B. Terhal, A no-go theorem for a two-dimensional self-correcting quantum memory based on stabilizer codes, New J. Phys. **11**, 043029 (2009).
[89] T. Rakovszky and V. Khemani, The physics of (good) LDPC codes II. Product constructions, ArXiv:2402.16831.
[90] J. Behrends and B. Beri, Statistical mechanical mapping and maximum-likelihood thresholds for the surface code under generic single-qubit coherent errors, ArXiv:2410.22436.
[91] Y. Bao and S. Anand, Phases of decodability in the surface code with unitary errors, ArXiv:2411.05785.
[92] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).
[93] S. Krastanov, F. Ahmad, P. Viswanathan, S. Pardis, A. Micciche, J. Lapeyre, Rabqubit, Gsommers, T. Hofmann, A. Bhatt, A. Meligrana, Benzillaist, C. Zhao, IsaacP1234, L. Göttgens, ShuGe-MIT, T. Holy, T. Dang, Adrianariton, and Ismoldayev, Quantumsavory/quantumclifford.jl: v0.9.12, 2024, https://doi.org/10.5281/zenodo.13950271.