

Observation of a fault tolerance threshold with concatenated codes

Grace M. Sommers^{1,*}, Michael Foss-Feig², David Hayes², David A. Huse¹, and Michael J. Gullans³

¹*Physics Department, Princeton University, Princeton, New Jersey 08544, USA*

²*Quantinuum, 303 S. Technology Ct., Broomfield, Colorado 80021, USA*

³*Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*



(Received 23 June 2025; accepted 27 October 2025; published 9 December 2025)

We introduce a fault-tolerant protocol for code concatenation of a generalized Shor code using a butterfly network architecture with high-noise thresholds and low ancilla overhead to allow implementation on current devices. We develop a probability passing decoder using tensor networks that applies Bayesian updates to the marginal error probabilities after each layer of checks, achieving a state-preparation threshold of $e_c \approx 0.089$ for erasure errors and ≈ 0.015 for unheralded noise. We implement our state-preparation protocol on ion-trap hardware with added noise to demonstrate the threshold behavior in a real quantum device. We further theoretically test the performance of our scheme as a quantum memory and for universal quantum computation through the preparation of low-noise magic states for state distillation and T -gate injection.

DOI: [10.1103/v477-jw8l](https://doi.org/10.1103/v477-jw8l)

I. INTRODUCTION

Recent years have seen dramatic experimental advances in quantum error correction and fault tolerance, ranging from demonstrations of fault-tolerant gadgets including repeated rounds of error correction [1–6], fault-tolerant algorithms with many logical qubits [7], and subthreshold scaling in topological codes [8]. Interestingly, however, fault-tolerant threshold behavior has not been experimentally observed in concatenated codes, despite their foundational role in fault-tolerant quantum computing [9–15].

At the same time, recent years have seen steady *theoretical* progress in the study of concatenated codes, from decoding algorithms [16–19], to ancilla preparation [20–22] and associated overhead analysis [23–27], to the construction of universal gate sets without magic state distillation [28–32]. More recently, there is an increasing interest in concatenated codes from a *practical* quantum computing perspective now that optimized schemes [33,34] have brought their performance (measured asymptotically by noise threshold and/or encoding rate) to a level on par with low-density parity check (LDPC) codes like the surface code [35] or high-rate codes on nonlocal geometries [36,37]. Moreover, from a statistical mechanics perspective, concatenated codes allow novel realizations of spin-glass phases and transitions in treelike geometries [16,18,38]. These developments motivate an investigation of the threshold behavior of concatenated codes on current quantum devices.

In this work, we introduce a family of efficient fault-tolerant protocols for code concatenation with high thresholds and low ancilla overheads to enable their realization on quantum hardware, at the expense of increased decoding complexity. We begin in Sec. II by introducing the concatenated code family, which consists of two-qubit repetition codes in alternating bases, and designing a state-preparation circuit that prepares the logical states $|0\rangle$ and $|\bar{\pm}\rangle$ with zero ancilla overhead. In Sec. III, we evaluate the state-preparation gadget's threshold performance under both heralded and unheralded noise in classical simulations and on the Quantinuum System Model H2, a trapped-ion quantum processor with all-to-all connectivity and midcircuit measurement [39]. The state-preparation gadget is then incorporated into a quantum memory test in Sec. IV, theoretically demonstrating the survival of logical information through many rounds of Steane error correction [40]. Finally, in Sec. V we analytically examine a protocol for preparing logical magic states below the distillation threshold [41], establishing a path toward universal quantum computation in our low-overhead and high-threshold concatenation scheme. We conclude in Sec. VI. The Appendices and Supplemental Material [42] contain a detailed exposition of the error models and decoding methods. The tensor-network-based approximate decoder introduced therein, which combines methods from belief propagation and weight enumerator tensor networks [43–45], may be of independent interest.

II. SINGLE- AND MULTITREE ENCODINGS

The encoding circuit of a self-concatenated code, with a base code on b physical qubits, can be written as a b -ary tree, whose depth t is the number of generations [18,38]. In a previous work [38], we investigated the phase diagrams of concatenated codes generated by binary ($b = 2$) trees; although the base code necessarily has distance one, the

*Contact author: gsommers@princeton.edu

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

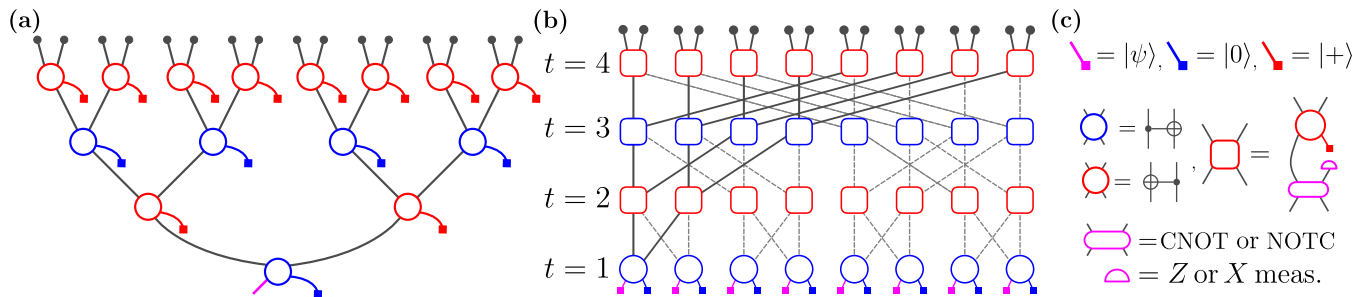


FIG. 1. (a) Example depth-4 encoding circuit of a GS_0 code. In each layer, the left qubit entering a gate comes from the previous layer, while the right qubit is in a fresh pure state. One logical qubit (purple leg at root) is encoded into 16 physical qubits (leaves of tree). (b) Multitree (butterfly network) state-preparation circuit for $|0\rangle$ (purple legs all $|0\rangle$) or $|+\rangle$ (purple legs all $|+\rangle$). Thick solid lines trace the tree that encodes the state on the “system,” corresponding to the branches of the singletree in panel (a), while dashed gray lines indicate qubits used as ancillas. (c) Legend for panels (a) and (b). Blue and red circles denote CNOT (left qubit as control) and NOTC (right qubit as control), respectively. Each square node of the butterfly network consists of (1) a “check gate” (CNOT for $|0\rangle$, NOTC for $|+\rangle$), (2) a measurement of the ancilla qubit in the corresponding basis (Z or X), used to read off the syndrome in that layer, (3) resetting the ancilla to a fresh stabilizer state, and (4) an encoding gate (CNOT for blue square nodes, NOTC for red square nodes). Each of these steps can incur errors, as described in the main text.

appropriate choice of encoding isometry can generate codes with distance growing exponentially in t . One of these is the binary tree version of a generalized Shor (GS) code [46,47], which alternatively concatenates a two-qubit repetition code in the Z basis and X basis. That is, as shown in Fig. 1(a) [48], the encoding isometry alternates between

$$\text{CNOT}_{12}|0\rangle_2 = |00\rangle\langle 0| + |11\rangle\langle 1| \quad (1)$$

and

$$\text{CNOT}_{21}|+\rangle_2 = |++\rangle\langle +| + |--\rangle\langle -|. \quad (2)$$

In the ensuing discussion, we will drop subscripts and denote $\text{NOTC} \equiv \text{CNOT}_{21}$. Where a distinction is meaningful, we use GS_0 (GS_e) to denote the code family where Eq. (1) occurs on odd (even) layers, mainly employing the former encoding order.

The $b=2$ GS code family has the parameters $[[n=2^t, k=1, d=2^{\lfloor t/2 \rfloor}]]$. While there exist other binary tree codes with superior distance scaling and performance under Pauli errors, the GS code is well suited to fault-tolerant protocols because it is a Calderbank-Shor-Steane (CSS) code, in which the logical operators and stabilizer generators can all be chosen to be all Z's or all X's [49,50]. CSS codes are automatically equipped with a transversal CNOT gate, and Steane error correction reduces the task of fault-tolerant syndrome measurement to the task of preparing ancillas encoded in $|0\rangle$ and $|+\rangle$ [40].

The binary tree encoding circuit offers one way to prepare a given logical state $|\bar{\psi}\rangle$: Feeding the single-qubit state $|\psi\rangle$ into the root of the tree, the output on the leaves is $|\bar{\psi}\rangle$. When $|\psi\rangle$ is a stabilizer state, stabilized by the Pauli operator P , we can think of this as setting the “logical leg” to be an additional stabilizer; e.g., to prepare the state $|0\rangle$, the root is stabilized by Z [51].

However, the encoding circuit inevitably suffers from noise, and errors near the root of the tree can accumulate and propagate through the circuit. Consequently, if we only read out the syndrome in the final layer, at any finite noise rate p , the logical failure probability $P_F(p, t)$ is monotone-increasing with tree depth t . Prior work [18,38] has demonstrated a threshold p_{single} below which *some* quantum information

survives to infinite depth, i.e., $\lim_{t \rightarrow \infty} P_F(p, t) < 1/2$. To make the logical failure rate converge to zero as $t \rightarrow \infty$, we could then perform state distillation [41], consuming many noisy states to produce one less noisy state.

The protocol we present here, for preparing the logical states $|0\rangle$ and $|+\rangle$ used in Steane error correction, improves upon that approach by introducing checks between each layer of encoding, rather than just at the end. This method achieves a higher threshold $p_c > p_{\text{single}}$ and requires zero qubit overhead, in the sense that the circuit preparing a depth- T encoded state requires only 2^T total qubits.

While interleaving checks in the encoding circuit is not a new idea—they are implicit in the hierarchical structure of many Steane EC-style ancilla preparation circuits [22], and a similar approach underlies the Knill-style Bell state-preparation protocol in Ref. [24]—they take an especially simple form for our code family, turning the singletree into a multitree known as a *butterfly network* [Fig. 1(b)] [52]. The same network geometry appears in the fast Fourier transform [53,54] and in the encoding circuit of a quantum polar code [55]. At level t , each depth- t state S is checked against a depth- t ancilla A , yielding partial information about the errors on the system up to that layer. The “checking” procedure for state preparation of $|0\rangle$ in the GS_0 code family consists of the following steps [Fig. 1(c)]:

- (1) Perform a transversal CNOT between S and A .
- (2) Measure each qubit of A in the Z basis. From these measurement outcomes, we can infer the Z syndrome of the joint system $S \otimes A$.
- (3) After odd (even) layers, reset each qubit of A to $|+\rangle$ ($|0\rangle$), and feed it in as a fresh stabilizer to the next level of encoding S , including the root stabilizers.

To prepare $|+\rangle$, we fix all the root stabilizers to be X, replace the transversal CNOTs with transversal NOTCs, and measure in the X basis. Therefore, at level t of preparing a depth- T state, there are 2^{T-t} identically prepared depth- t states. After the T th level of encoding, there is only one state left. As is standard in fault tolerance, to analyze the performance of this gadget in isolation, we conclude the circuit by perfectly measuring all the stabilizers, *except* the root stabilizer. The decoder then uses the syndromes measured

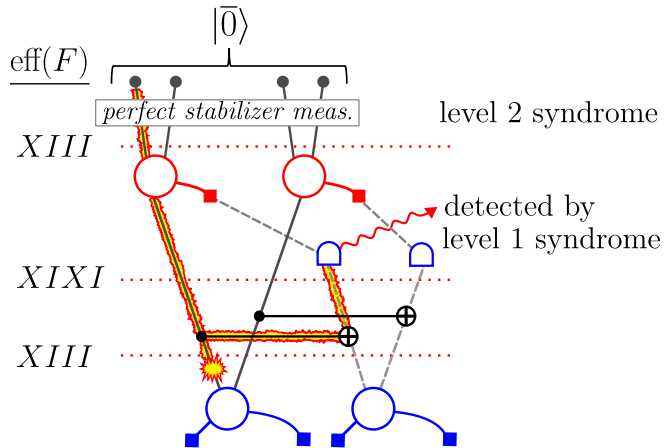


FIG. 2. Depth-2 multitree circuit for $|\bar{0}\rangle$, with a round of perfect stabilizer measurements appended to evaluate its performance. Highlighted wires show how a fault (yellow starburst) propagates through the circuit. The first round of checks (level 1 syndrome) distinguishes this fault from a bit flip on the second system qubit, which would propagate to a logically inequivalent error with the same level 2 syndrome.

in levels 1, 2, ..., T , which collectively form the “spacetime syndrome,” to (1) find a Pauli error F in spacetime that produces this syndrome and (2) decide whether to apply $\text{eff}(F)$ or $\text{eff}(F)L$. Here, $\text{eff}(F)$ is the fault propagated to the last layer of the circuit [56] and L is the logical operator that flips the root stabilizer ($L = \bar{X}$ for $|\bar{0}\rangle$, $L = \bar{Z}$ for $|\bar{+}\rangle$). As a minimal example, Fig. 2 shows a fault with $\text{eff}(F) = X_1$. This fault would be decoded incorrectly if we only measured the syndrome in the final layer, but is correctly decoded using the full spacetime syndrome.

In previous overhead analysis of fault-tolerant protocols for concatenated codes [27], each round of stabilizer checks consists of both X -type and Z -type checks. In contrast, our protocol only performs one type of check, enabling zero qubit overhead as each ancilla is eventually reset and fed into the growing system. In the context of state preparation, this is justified because the decoder succeeds as long as it correctly fixes the sign of the root stabilizer (returns it to $+1$). Therefore at each level, we need only measure the syndrome that “matches” the root stabilizer, which for $|\bar{0}\rangle$ is all Z ’s (and thus immune to phase flips) and for $|\bar{+}\rangle$ is all X ’s (and thus immune to bit flips). This strategy is modified when using the prepared states for Steane error correction, discussed below.

III. STATE-PREPARATION THRESHOLDS

A. Theoretical demonstration

To test the protocol introduced in the previous section, we first consider a theoretical noise model with Pauli errors in three different locations: (1) on each stabilizer input (that is, on all the input legs before $t = 1$, and after every qubit reset in subsequent steps), (2) after each two-qubit gate, and (3) before each measurement on the ancilla (see Appendix A) [57].

1. Erasure errors

We first consider a model in which all errors are *heralded*: Each error raises a flag on the noisy qubit, while unflagged qubits are guaranteed to have suffered no errors. The most common heralded error channel is the erasure channel, in which the flagged qubit is replaced with a fully mixed state:

$$\mathcal{E}_e(\rho) = (1 - e)\rho \otimes |0\rangle\langle 0| + e\frac{\mathbb{I}}{2} \otimes |1\rangle\langle 1|, \quad (3)$$

where e is the erasure rate and $|0\rangle, |1\rangle$ is the state of a classical register. For CSS codes, the erasure channel is equivalent to the composition of a heralded bit flip channel

$$\mathcal{E}_{X,e}(\rho) = (1 - e)\rho \otimes |0\rangle\langle 0| + \frac{e}{2}(\rho + X\rho X^\dagger) \otimes |1\rangle\langle 1|, \quad (4)$$

with a heralded phase flip channel $\mathcal{E}_{Z,e}$.

From the circuit architecture, we construct the stabilizer generators of the spacetime code [56,58,59], wherein each possible error location is assigned a qubit, and each bit of syndrome information is associated with a stabilizer generator in spacetime. Under fully heralded noise, optimal decoding of this spacetime code can be done in polynomial time (see Sec. II of the Supplemental Material [42]). For independent and identically distributed (i.i.d.) erasures at rate e , the threshold for both $|\bar{+}\rangle$ and $|\bar{0}\rangle$ is $e_c \approx 0.089$, as shown in Fig. 3(a)(i). The crossing between system sizes drifts to smaller e and smaller $P_F(e)$ with increasing T . A fit to the one-parameter scaling form

$$P_F(e, T) = f((e - e_c)N(T)^{1/\nu}), \quad (5)$$

with f an unknown scaling function, $e_c = 0.089$, $\nu = 2.85$, and $N(T) \sim T2^T$ the number of spacetime qubits, is shown in Fig. 3(a)(ii) [60].

2. Stacked probability passing

Next, we consider unheralded noise, taken to be independent bit and phase flips at rate r :

$$\mathcal{N}_r(\rho) = (\mathcal{N}_{X,r} \circ \mathcal{N}_{Z,r})(\rho), \quad (6)$$

where

$$\mathcal{N}_{X,r} = (1 - r)\rho + rX\rho X^\dagger, \quad (7a)$$

$$\mathcal{N}_{Z,r} = (1 - r)\rho + rZ\rho Z^\dagger. \quad (7b)$$

The lack of correlations between X and Z errors means that, for CSS codes, we can decode the Z and X syndromes independently.

Without perfect heralding, optimal decoding via the spacetime code is no longer feasible. Instead, we use an approximate “stacked probability passing” decoder, detailed in Appendix B and the Supplemental Material [42]. Under this noise model, the stacked probability passing decoder has an estimated threshold of $r_c \approx 0.015$, as shown in Fig. 3(b).

To benchmark the approximate decoder, we also use it to decode erasure errors, comparing its performance to that of the optimal decoder. For that error model, the approximate decoder achieves a similar threshold and subthreshold performance to the optimal decoder, though further systematic improvements could be made [42].

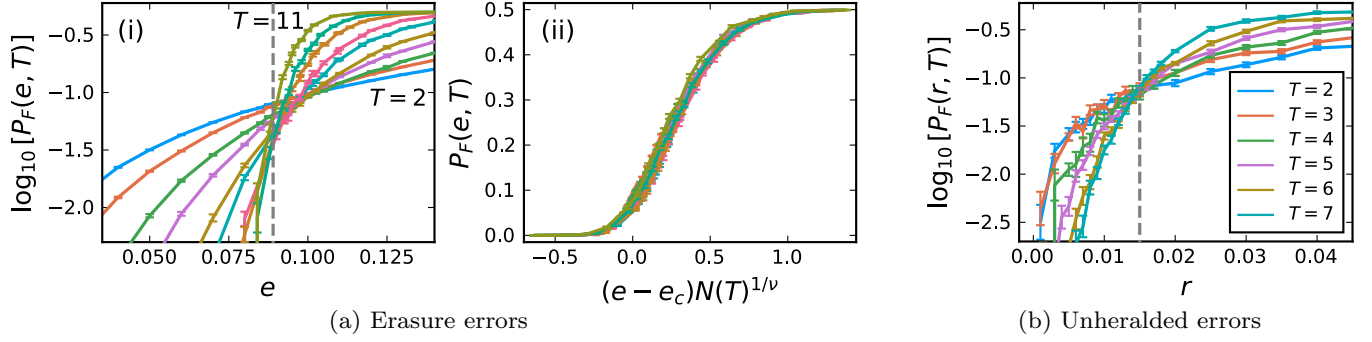


FIG. 3. Logical failure probability of state preparation in classical simulations. (a) Erasure errors at rate e , optimal decoder. Left panel (i) shows system sizes $T = 2$ to $T = 11$, plotted on a log scale. Gray dashed line marks the threshold $e_c \approx 0.089$. Right panel (ii) is a scaling collapse to the form (5) with $e_c = 0.089$, $\nu = 2.85$, plotted on a linear scale with depths $T = 3-11$. (b) Unheralded bit/phase flips at rate r , decoded via stacked probability passing. Gray dashed line marks the threshold $r_c \approx 0.015$. In all panels, the plotted quantity is averaged over $|\bar{0}\rangle$ and $|\bar{\oplus}\rangle$. In this paper, all logarithms are base 10, and uncertainties are determined as the standard error of the mean across independent samples or the one sigma confidence interval across Bernoulli trials. We use a linear scale for the horizontal axis as our focus is on the near-threshold behavior, not the behavior near $p = 0$.

B. Hardware implementation

Having demonstrated the effectiveness of our circuit layout and decoding method in theory, we now test the protocol on quantum hardware.

We implement state-preparation circuits using up to 32 qubits (depth $T = 5$) on the Quantinuum System Model H2. The dominant error sources native to the device are two-qubit gate errors and measurement errors, each occurring at a rate $\lesssim 0.0025$. Modeled as independent bit and phase flips, this noise rate is well below r_c . To tune through a transition, we stochastically introduce gates in two different ways: (1) incoherent noise and (2) coherent rotations. In the final layer of circuit, each qubit in the system is destructively measured in the basis of the root stabilizer; decoding is successful if the decoder's guess of the logical state, using the heralding record and syndrome, matches the measured state.

We choose both types of added errors to be heralded, so the threshold is at a large enough added noise rate to clearly distinguish the subthreshold and above-threshold phases.

1. Added incoherent errors

The incoherent error model is heralded bit flips [4] (for $|\bar{0}\rangle$) and phase flips (for $|\bar{\oplus}\rangle$) at rate e after each two-qubit gate and before each measurement. Even at these modest system sizes, Fig. 4(a) shows evidence of a threshold between $e = 0.06$ and $e = 0.09$. In the subthreshold phase, the logical failure probability is exponentially suppressed in T , as shown in the inset.

We stress that while the form of the added noise is artificial, the exponential suppression of the logical failure probability at low rates of added noise already demonstrates the success of the protocol in handling the noise native to the device. Injecting noise of our choosing increases the logical failure rate, allowing the trend with system size to be resolved using fewer shots.

2. Added coherent errors

We next benchmark the state-preparation protocol in the presence of added coherent errors at rate q . That is, we

introduce coherent non-Clifford rotations (with perfect heralding to boost the threshold): If a flag is raised (with probability q), then a coherent rotation by θ about the X (or Z) axis on the Bloch sphere is applied:

$$\mathcal{C}_{X,q,\theta}(\rho) = (1-q)\rho \otimes |0\rangle\langle 0| + qe^{i\theta X/2}\rho e^{-i\theta X/2} \otimes |1\rangle\langle 1|. \quad (8)$$

We set the rotation angle $\theta = \pi/4$, so that the rotation is maximally non-Clifford.

Our motivation for considering this second model is that while incoherent noise is typically assumed when developing decoders, real devices also suffer systematic small rotations [61]. Part of the difficulty in studying coherent noise is the phenomenon of quantum interference, which makes syndrome sampling classically inefficient in general and gives rise to complex Boltzmann weights in the associated statistical mechanics mapping [62]. With quantum hardware, the inefficiency of syndrome sampling is circumvented as the syndromes are generated for us by the circuit. Though the heralding is unphysical, it serves only to boost the threshold and does not alter the implementation of the decoder.

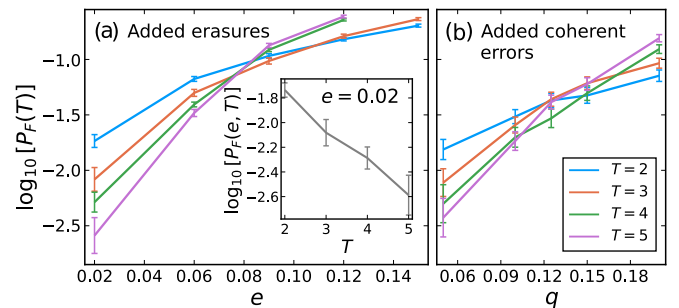


FIG. 4. Logarithm of the failure probability on Quantinuum System Model H2, native noise + added heralded bit/phase flips at rate e [Eq. (4)] (a) and heralded coherent errors [Eq. (8)] at rate q (b). The inset to panel (a) shows the subthreshold scaling at added noise rate $e = 0.02$. In all panels, the plotted quantity is averaged over $|\bar{0}\rangle$ and $|\bar{\oplus}\rangle$.

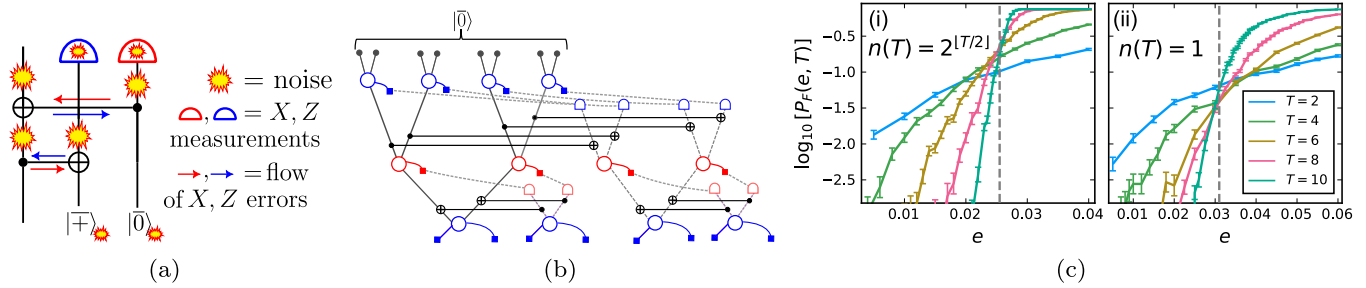


FIG. 5. (a) Steane syndrome measurement. Leftmost wire is an unknown logical state on the system. Yellow starbursts show the noise locations modeled in this work. Arrows indicate the flow of Pauli errors. (b) Depth-3 preparation circuit for an ancilla encoded in $|\bar{0}\rangle$, for use in the gadget shown in panel (a). This preparation circuit includes X checks after the first layer and Z checks after the second layer. Legend is the same as in Fig. 1(c). (c) Logical failure probability after (i) $2^{\lfloor T/2 \rfloor}$ rounds and (ii) 1 round of Steane syndrome measurement, under a bulk erasure model with optimal decoding. Apparent thresholds, achieving the best collapse to the scaling form (5), are indicated by dashed gray lines. Uncertainties are determined as the standard error of the sample mean.

Although the added errors are coherent, our approximate decoder still models the noise as incoherent by taking the Pauli twirl,

$$\tilde{\mathcal{C}}_{X,q,\theta}(\rho) = q[\cos^2(\theta/2)\rho + \sin^2(\theta/2)X\rho X^\dagger] \otimes |1\rangle\langle 1| + (1-q)\rho \otimes |0\rangle\langle 0|. \quad (9)$$

At small q , this decoder achieves modest suppression of the failure probability with increasing T , with an apparent threshold around $q \approx 0.125$ [Fig. 4(b)].

3. Future directions

A potential optimization, enabled by the hardware's capacity for midcircuit measurement and feedback, is to adapt the checking procedure in real time based on syndrome information from earlier rounds. While a rewiring protocol for a fully heralded error model does not yield asymptotic improvements for preparation of $|\bar{0}\rangle$ and $|\bar{+}\rangle$, adaptive rewiring may be an asset for other tasks [42].

Successful preparation of $|\bar{0}\rangle$ and $|\bar{+}\rangle$ only demonstrates the code's *classical* performance, since the measurements are performed on a single basis. As a more stringent test of the code performance, one could prepare a logical Bell pair, which would require defense against both types of noise simultaneously. In an experiment, the error-corrected correlation between measurement outcomes would serve as a measure of entanglement in the logical state [63]. However, the logical Bell state uses twice as many qubits for a given T as $|\bar{0}\rangle$ and $|\bar{+}\rangle$, thus limiting the number of generations accessible on current hardware. We therefore opt not to pursue this avenue, instead turning to a *theoretical* demonstration of the code's use as a quantum memory.

IV. QUANTUM MEMORY

For the quantum memory demonstration, we use the encoded logical states as ancillas in a Steane syndrome extraction gadget. A review of Steane error correction and details of our implementation are provided in Sec. VI of the Supplemental Material [42]. For the present discussion, it suffices to recall [see Fig. 5(a)] that one round of Steane error correction uses two ancillas, encoded in the states $|\bar{0}\rangle$ and $|\bar{+}\rangle$ and coupled to the system via transversal CNOT gates, to

perform X and Z checks on a CSS code *without* learning the logical state [40]. These ancillas are prepared via a modification of the state-preparation protocol discussed above. The state-preparation circuit in Fig. 1(b) defends against only one type of errors; for example, the circuit for $|\bar{0}\rangle$ uses Z checks in each layer, allowing phase flip errors to accumulate since they do not affect the logical state. In contrast, when $|\bar{0}\rangle$ is used in the Steane EC gadget, we must defend against both bit flip errors (which otherwise will leak onto the system during the transversal CNOT) and phase flip errors (since $|\bar{0}\rangle$ is used to learn the X syndrome). To address this issue while maintaining the butterfly network structure of the circuit, we simply alternate between X and Z checks in the multitree, as shown for a depth-3 ancilla in Fig. 5(b).

Suppose that an unknown logical state has been perfectly encoded in a depth- T tree code. We can think of this state as being entangled with a reference qubit. We then use $n \geq 1$ rounds of the syndrome extraction gadget to read out the syndrome. Since the ancillas used in the syndrome extraction are noisy, and the system can suffer additional errors after each layer of gates, the decoder's task is to predict the correction operator that returns the system to its initial (unknown) state, based on the syndrome information read out from the ancilla measurements and from the perfect stabilizer measurements at the end of the circuit.

When the errors are heralded, we infer the performance of an optimal decoder by computing the mutual information between the final state of the system S and the initial reference R . Perfect decoding is possible when the mutual information is maximal, $I(S : R) = 2$. Otherwise, for each lost bit of mutual information, the recovery probability degrades by a factor of 2, such that the logical failure probability of the optimal decoder for a given $I(S : R)$ is

$$P_F(e) = 1 - \frac{1}{2^{2-I(S:R)}}. \quad (10)$$

To test the survival of logical information after many rounds of syndrome extraction, we subject a depth- T logical state to $n = d(T) = 2^{\lfloor T/2 \rfloor}$ rounds of noisy syndrome measurement, followed by a layer of perfect stabilizer measurements. To test the performance of a single Steane error correction gadget in isolation, we also examine the logical failure probability for $n = 1$ round. The thresholds for the two circuits are

$e_c \approx 0.0255$ and $e_c \approx 0.031$, respectively, as shown in Fig. 5. Both transitions are broadened with an exponent $\nu \approx 2.75$, with “system size” $N(T) \sim T^{2.75} n(T)$.

The approximate probability of passing decoder described in Sec. VII of the Supplemental Material [42] has a much lower threshold and inferior subthreshold scaling. Therefore, a transition under unheralded errors is only visible if we restrict to $n = 1$ rounds of checks. In this case, we attain a pseudothreshold of $r_c \approx 0.003$ below which the failure probability does not increase with depth. We leave improvements of the decoder for unheralded and partially heralded noise to future work.

V. UNIVERSAL GATES

Finally, we theoretically examine how to construct a fault-tolerant universal gate set for the GS code family. For this purpose, it suffices to implement the generators of the Clifford group—CNOT, Hadamard, and the phase gate S —supplemented with a logical T gate.

As a CSS code, the GS code automatically comes equipped with a transversal CNOT, a fact exploited above in the state-preparation and Steane-EC gadgets. We propose to implement a logical Hadamard using a transversal gate + code switching [64–70], discussed below. To implement the remaining gates at the logical level, we rely on gate injection/gate teleportation [41,71]: (1) prepare an ancilla in a logical resource state, (2) couple the ancilla to the system via transversal CNOT, (3) measure the ancilla, and (4) conditioned on the measurement outcome, apply an (easier-to-implement) gate on the system.

A. Magic state distillation

We first consider state preparation of logical T states, which can be used for T -gate injection [41]. Specifically, we prepare the state [72]

$$|\bar{T}\rangle \equiv e^{-i\pi\bar{Z}/8}|\bar{+}\rangle. \quad (11)$$

Since this state does not point along the X or Z axis of the logical Bloch sphere, it cannot be prepared using the multitree circuit in Fig. 1(b). Instead, we prepare the state using the singletree circuit [Fig. 1(a) with $|T\rangle = \exp(-i\pi Z/8)|+\rangle$ fed into the root], followed by state distillation.

Let $\mathcal{N}$ denote the noisy encoding isometry (from the state $|T\rangle$ on the root to a noisy encoded state on the leaves) and $\mathcal{R}$ the optimal recovery channel. If the noise + decoder implements a logical P_α with probability $\mathbb{P}_\alpha$, then the logical T state fidelity is

$$F(\bar{T}) \equiv \langle \bar{T} | \mathcal{R} \circ \mathcal{N}(|T\rangle\langle T|) | \bar{T} \rangle = 1 - \mathbb{P}_Z - (\mathbb{P}_X + \mathbb{P}_Y)/2. \quad (12)$$

For models of bulk errors parameterized by p , with perfect syndrome measurement in the final layer, there exists a “coding phase” in which $F(\bar{T})$ converges, in the limit of infinite depth, to a fixed point $F^*(p) > 1/2$ for $p < p_{\text{single}}$. If the noise is composed of single-qubit independent unheralded bit and phase flips, the threshold can be determined numerically to high precision as $r_{\text{single}} = 0.0066 \pm 0.0004$, whereas under erasure errors, an exact calculation yields $e_{\text{single}} = 0.05505\dots$

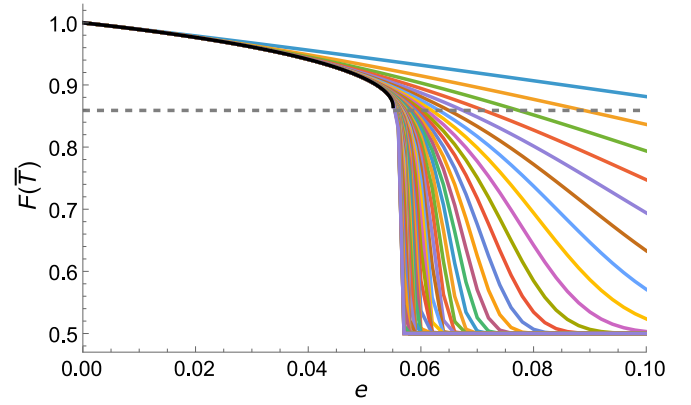


FIG. 6. Fidelity to the logical T state [Eq. (12)] after decoding erasure errors on the branches of the singletree encoding circuit. Colored curves are for even depths up to 100, using the GS_e encoding order [74]. Black curve shows the fixed-point fidelity. Dashed line shows the threshold for magic state distillation from Ref. [41].

[38]. Figure 6 shows the fidelity as a function of erasure rate e , for even depths up to 100. The fixed-point fidelity within the coding phase is marked as a black curve.

Optimal decoding of the singletree is followed by distillation. To converge to zero failure probability, the fidelity of the inputs to the distillation protocol must exceed a certain value, which equals 0.859 in the 15-to-1 scheme introduced in Ref. [41]. An analytic calculation for erasure errors finds that for all $e \leq e_{\text{single}}$, $F^*(e) > 0.859$, so distillation is successful throughout the entire coding phase [73]. Additional details are provided in Sec. III C of the Supplemental Material [42].

B. Logical Clifford gates

The standard T -gate injection circuit uses a logical phase gate S , conditioned on the measurement outcome on the ancilla [71]. In turn, we can also implement S fault-tolerantly using gate injection, with the resource state $|\bar{i}\rangle = (|\bar{0}\rangle + i|\bar{1}\rangle)/\sqrt{2}$, the unique $+1$ eigenstate of $\bar{Y}$. To prepare $|\bar{i}\rangle$, we once again use the singletree encoding circuit, now with the logical leg fixed to be a $-Y$ stabilizer, followed by a distillation procedure that produces one less noisy $|\bar{i}\rangle$ from seven noisy $|\bar{i}\rangle$'s [25]. The classically conditioned gate in step 4 of the gate injection circuit is just a Pauli gate, which admits a transversal implementation.

To complete our universal gate set, we need a fault-tolerant Hadamard gate. Applying H transversally does not quite implement a logical Hadamard: It also exchanges the X -type stabilizers with Z -type stabilizers, switching between the GS_o and GS_e code family. We can account for this by simply keeping track of which of the two codes each logical qubit is in. The only complication is that before performing a gate between a GS_o -encoded qubit and a GS_e -encoded qubit, we must *code switch* one of the two qubits to put them in the same code.

The code-switching protocol works as follows. Suppose we want to code switch from $\text{GS}_e(t)$ to $\text{GS}_o(t)$. To do so, we need to project onto the codespace of GS_o , while preserving the logical state [70]. Projecting onto the codespace is simple: just measure the syndrome using GS_o -encoded ancillas, and fix up the signs of the wrong-signed stabilizers by applying

Pauli corrections. As long as these Pauli corrections commute with the logical operators, the logical state is preserved. To meet this condition, we first permute the physical qubits, mapping (for even t) the time-evolved operators ($X_e(t)$, $Z_e(t)$), which are logical representatives of the $\text{GS}_e(t)$ code, onto the operators ($X_o(t)$, $Z_o(t)$) of $\text{GS}_o(t)$. Then, the permuted code $\widetilde{\text{GS}}_e(t)$ and the target code $\text{GS}_o(t)$ are different gauge fixings of a common subsystem code [66], so the Pauli corrections can be implemented using stabilizers of $\widetilde{\text{GS}}_e$, which preserve the logical operators. Further detail is provided in Sec. VIII of the Supplemental Material [42].

VI. DISCUSSION

In this work, we have taken steps toward universal fault-tolerant quantum computation using a family of concatenated CSS codes whose base code consists of two qubits. We constructed optimal and approximate decoders for a zero-ancilla-overhead state-preparation protocol, and demonstrated thresholds on trapped-ion hardware. The state-preparation circuit was then adapted for use in a quantum memory test. Finally, we showed how the bare (singletree) circuit, followed by state distillation, can be used to prepare resource states such as the logical T state, an essential ingredient for performing universal logical gates.

Concatenated codes with Steane error correction, and LDPC codes such as the surface code, are two paths toward fault-tolerant quantum computing. One advantage of the former approach is that clever choices of concatenation enable fault-tolerant implementation of universal gates without gate injection, thus circumventing magic state distillation (or cultivation [75]) entirely [27–32]. While Ref. [27] found that such distillation-free concatenated codes still require more resources to achieve the same logical error rate as surface codes with magic state distillation, their approach is dominated by the overhead associated with the preparation and verification of ancilla states for Steane error correction, which we apply more sparingly. In particular, our state-preparation circuit has zero qubit overhead, in contrast with many previously introduced protocols that rely on postselection, incurring significant time and qubit overhead [13,20,23,40]. The GS code used for our demonstration *does* require magic states to perform universal gates, but the high-level strategy outlined in this work—multitree state preparation with qubit reuse and no postselection, made possible by a tensor-network-based decoder—applies to general concatenated CSS codes. The simplest way to generalize the multitree circuit to base codes with $b > 2$ would be to replace Fig. 1(b) with a radix- b butterfly network, meaning that each layer of concatenation would be followed by checking a system block against $b - 1$ ancilla blocks, whose qubits are then reset and fed into the next layer. A fault occurring in the last layer can only spread to at most b qubits, which is a bounded error in a code with macroscopic distance ($t \rightarrow \infty$, b fixed).

The appeal of the GS code lies in the simplicity of one concatenation level, which involves just a single two-qubit gate in contrast to the larger base codes studied in earlier works. With large base codes, analysis is often limited to one or two concatenation levels, yet the “pseudothreshold” inferred from just one concatenation level can differ significantly from the

asymptotic threshold [29,76,77]. With our small base code, we can access multiple generations on the hardware, demonstrating a suppression of the logical failure probability in the subthreshold phase, and many more generations in classical simulations, enabling a finite-size scaling analysis.

It is worth noting that while our multitree state preparation avoids the *qubit* overhead involved in a protocol that applies postselection or distillation of several singletrees, it incurs a *decoder* overhead due to the relative complexity of the stacked probability passing decoder vs the singletree decoder. Decoding is more efficient, and the threshold is much higher, when the errors are heralded. While perfect heralding of errors is an idealized scenario, qubits on a range of experimental platforms can be designed so that *most* errors are erasures, via a technique known as erasure conversion [78–89]. A preliminary investigation indicates that the threshold of the state-preparation gadget increases sharply as the heralding rate approaches 1, consistent with the findings of Ref. [78] for the surface code.

Finally, as noted above, the butterfly network geometry appears in several other contexts, and an avenue for future study is to consider how extensions in those contexts translate into generalizations of the multitree state-preparation circuit. For example, mixed-radix algorithms for the fast Fourier transform when the input size is not a power of 2 [90] could be useful for preparing codes with different base codes in different layers. The encoding circuit of a quantum polar code is also a butterfly network, with each node corresponding to a CNOT gate rather than the “check and reset” black box in Fig. 1(c). As polar codes are a special case of branching MERA codes [55], it may be interesting to explore the circuit that results when each node of the branching MERA graph is replaced with a check and reset.

ACKNOWLEDGMENTS

We thank Zhi-Yuan Wei for pointing out the connection to quantum polar codes and branching MERA. This research was supported in part by NSF QLCI Grant No. OMA-2120757, including an Institute for Robust Quantum Simulation (RQS) seed grant. We gratefully acknowledge the Quantum Computing User Program (QCUP) at Oak Ridge National Lab for providing the credits for running on the Quantinuum device. Numerical work was completed using computational resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology’s High Performance Computing Center and Visualization Laboratory at Princeton University. The open-source QuantumClifford.jl package was used to simulate Clifford circuits with erasure errors [91]. Any mention of commercial products in this paper is for information only; it does not imply recommendation or endorsement by NIST, nor does it imply that the equipment identified is necessarily the best available for the purpose.

DATA AVAILABILITY

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

APPENDIX A: ERROR MODELS

In this Appendix, we summarize the various error models considered in the work.

1. Types of errors

Three types of noise channels are considered in this work: erasure errors [Eq. (3)], which in a CSS code are equivalent to independent heralded bit and phase flips $\mathcal{E}_{X,Z}$ [Eq. (4)]; independent unheralded bit and phase flips $\mathcal{N}_{X,Z}$ [Eq. (6)]; and heralded coherent errors $\mathcal{C}_{X,Z}$ [Eq. (8)]. To emphasize the difference between the channels $\mathcal{E}$, $\mathcal{N}$, $\mathcal{C}$, we parametrize them by different letters: r , e , and q , respectively [93]. When not referring to a specific channel, we parametrize the noise rate by p .

In the state-preparation gadget, noise can be treated classical, because the logical state $|\bar{0}\rangle$ ($|\bar{+}\rangle$) is immune to Z (X) errors. Thus, we need only consider $\mathcal{E}_X$, $\mathcal{N}_X$, $\mathcal{C}_{X,\pi/4}$ in preparing $|\bar{0}\rangle$ and vice versa for $|\bar{+}\rangle$.

2. Error locations

All the error models considered in this work consist of independent and identically distributed single-qubit errors, at three different types of locations:

- (1) On each stabilizer input, at rate p_i .
- (2) On the two output wires of each two-qubit gate, at rate p_g .
- (3) Before each measurement (i.e., random bit flip of measurement outcome), at rate p_m .

In classical simulations [Figs. 3(a) and 5], we take $p_i = p_g = p_m = p$, though gate errors have the largest effect on the performance due to the larger number of available error locations and the potential for errors to spread in subsequent layers.

In the experiment, we first infer the native unheralded noise rates r_i , r_g , r_m from the reported rate of two-qubit gate errors and state preparation and measurement (SPAM) errors. At the time of the experiment, $p_{\text{SPAM}} \approx 0.0025$ and is dominated by measurement errors, so we take $r_i = 0$, $r_m = 0.0025$. The probability of a nontrivial two-qubit Pauli operator following a two-qubit gate was $p_{2qb} \lesssim 0.002$. Our decoder models these errors as independent X and Z errors on each qubit [Eq. (6)], i.e., $1 - p_{2qb} = (1 - r_g)^4$, yielding $r_g \approx 0.0005$.

Then, we introduce either (1) heralded bit or phase flips at rate e or (2) heralded coherent errors at rate q , after each gate. For (1), we also add heralded bit flips of the measurement outcomes in postprocessing.

In the logical T state preparation, we consider noiseless measurements on the leaves and single-qubit erasure errors on every branch of the tree, including the root.

APPENDIX B: DECODING METHODS

Consider a faulty circuit (for state preparation or syndrome extraction) with several noisy measurement layers. To assess the performance of this gadget in isolation, we include a layer of perfect stabilizer measurements at the end of the circuit. The decoding task then consists of the following steps:

- (1) Given the syndrome information collected in each layer of measurements, find a Pauli operator E that, when

applied to the system qubits at the end of the circuit, returns us to the code space.

- (2) Determine which of the logically inequivalent classes is more likely.

(i) For state preparation, there are only two logical classes to consider, EL and E , where L is the logical operator that flips the logical state being prepared ($\bar{X}$ for $|\bar{0}\rangle$, $\bar{Z}$ for $|\bar{+}\rangle$).

(ii) For Steane syndrome extraction, there are four logical classes, but if X and Z errors occur independently, the decoder independently chooses from $\{\bar{I}, \bar{X}\}$, $\{\bar{I}, \bar{Z}\}$.

- (3) Apply the operator EL^* , where L^* is the most likely logical class in step (2). Decoding fails if the actual fault that occurred propagates to an error that differs from EL^* by a nontrivial logical operator.

An optimal maximum likelihood decoder computes the probability of each logical class in step (2) exactly. A sub-optimal decoder approximates the likelihood of each logical class, for example, by using a simplified error model (as when we use the Pauli twirl approximation for coherent errors, or model the native noise as independent bit and phase flips) or by discarding some syndrome information.

1. Erasure errors

When the noise is perfectly heralded, we can decode optimally. In theory, this is achieved by mapping the faulty circuit to a spacetime code, which is a subsystem code described further in Sec. II of the Supplemental Material [42]. In practice, the logical failure probability of this decoder can be determined without implementing the actual decoder, by simulating the noisy circuit evolution in a stabilizer tableau. The simulation is efficient because the erasure operation—tracing out a qubit—can be done in the stabilizer formalism, with worst-case complexity $O(N^2)$. In the state-preparation protocol, each root stabilizer is set to Z or X , and the entropy of the output state $S(\rho_S)$ is computed after the final round of perfect stabilizer measurements. The decoder succeeds with probability $1/2^{S(\rho_S)}$, where $S(\rho_S)$ is either 0 or 1, so in a given shot i , $P_F(i) = S(\rho_S(i))/2$. This quantity is then averaged over N_{shots} shots, separately for $|\bar{0}\rangle$ and $|\bar{+}\rangle$, to obtain $P_F(|\bar{0}\rangle)$, $P_F(|\bar{+}\rangle)$. In Fig. 3(a)(ii), we plot

$$P_F = \frac{1}{2} [P_F(|\bar{0}\rangle) + P_F(|\bar{+}\rangle)], \quad (\text{B1})$$

with error bars

$$\sqrt{[\text{SE}(P_F(|\bar{0}\rangle))]^2 + [\text{SE}(P_F(|\bar{+}\rangle))]^2} / 2, \quad (\text{B2})$$

where $\text{SE}(y) = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N_{\text{shots}}(N_{\text{shots}} - 1)}}$ is the standard error of the sample mean of y . In Fig. 3(a)(i) and figures in the Supplemental Material [42], to better discern the subthreshold performance, we plot the average of the log failure probability,

$$\log_{10}[P_F] = \frac{1}{2} [\log_{10}(P_F(|\bar{0}\rangle)) + \log_{10}(P_F(|\bar{+}\rangle))]. \quad (\text{B3})$$

In the quantum memory demonstration, the input state is a logical Bell pair $(|\bar{0}\rangle_S |0\rangle_R + |\bar{1}\rangle_S |1\rangle_R) / \sqrt{2}$, and the decoder succeeds with probability $1/2^{2-I(S:R)}$ where $I(S:R)$ is the mutual information between the system and reference [cf. Eq. (10)].

2. Probability passing

In the general case, the evaluation of logical class probabilities can be implemented via the contraction of a tensor network.

In the singletree setting used to prepare logical T states, with measurements only performed in the final layer, this tensor network has the same tree structure as the circuit itself [Fig. 1(a)]. The tree has an open leg at the root, and an open leg sticking out of each possible error location in the bulk. Since faults at different locations occur independently, a vector of error probabilities can be contracted locally on each bulk leg, yielding a tree tensor network whose output (on the root leg) is the vector of logical class probabilities.

In the fault-tolerant multitree setting, each layer of ancilla measurements gives us partial information about the errors occurring on the system. With this information, we can update the distribution of errors on the system tree via Bayes' rule. Optimal decoding would entail updating the full distribution exactly, introducing correlations between errors at different locations. The resulting tensor network is no longer efficiently contractible due to the large entanglement.

We therefore instead implement an approximate decoder that uses the idea of *probability passing*. A simple approxima-

tion would ignore all correlations by applying the Bayesian update marginally on each error location. For state preparation, this method achieves a pseudothreshold below which the logical failure probability is roughly independent of system size. To achieve a subthreshold phase in which the logical failure probability decreases with system size, our “stacked probability passing decoder” improves upon this crude approximation by keeping the correlations that develop from ancillas that interact directly with the system. The tensor network implementation of this algorithm is detailed in Sec. IV of the Supplemental Material [42].

Each shot of the state-preparation circuit yields a sample $y_i \in \{0, 1\}$, where $y = 0$ denotes success and $y = 1$ denotes failure. The average failure probability for preparing state $|\bar{\psi}\rangle = |\bar{0}\rangle, |\bar{+}\rangle$ is then

$$P_F(|\bar{\psi}\rangle) = \frac{\sum_i y_i}{N_{\text{shots}}}, \quad (\text{B4})$$

with estimated one sigma confidence interval $\sigma \approx \sqrt{P_F(1 - P_F)/N_{\text{shots}}}$. The uncertainty on $\log_{10}[P_F(|\bar{\psi}\rangle)]$ and its average across $|\bar{0}\rangle, |\bar{+}\rangle$ is determined through error propagation.

-
- [1] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, *Nature (London)* **598**, 281 (2021).
 - [2] L. Postler, F. Butt, I. Pogorelov, C. D. Marciniak, S. Heußen, R. Blatt, P. Schindler, M. Rispler, M. Müller, and T. Monz, Demonstration of fault-tolerant Steane quantum error correction, *PRX Quantum* **5**, 030326 (2024).
 - [3] S. Huang, K. R. Brown, and M. Cetina, Comparing Shor and Steane error correction using the Bacon-Shor code, *Science Advances* **10**, eadp2008 (2024).
 - [4] A. Paetznick, M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, J. P. C. III, A. Chernoguzov, J. M. Dreiling, C. Foltz, F. Frachon, J. P. Gaebler *et al.*, Demonstration of logical qubits and repeated error correction with better-than-physical error rates, [arXiv:2404.02280](https://arxiv.org/abs/2404.02280).
 - [5] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G. J. Norris, C. K. Andersen, M. Müller, A. Blais, C. Eichler, and A. Wallraff, Realizing repeated quantum error correction in a distance-three surface code, *Nature (London)* **605**, 669 (2022).
 - [6] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, *Nature (London)* **614**, 676 (2023).
 - [7] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter *et al.*, Logical quantum processor based on reconfigurable atom arrays, *Nature (London)* **626**, 58 (2024).
 - [8] R. Acharya, D. A. Abanin, L. Aghababae-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev *et al.*, Quantum error correction below the surface code threshold, *Nature (London)* **638**, 920 (2025).
 - [9] P. W. Shor, Fault-tolerant quantum computation, in *Proceedings of the 37th Conference on Foundations of Computer Science* (IEEE, Burlington, VT, 1996), pp. 56–65.
 - [10] E. Knill and R. Laflamme, Concatenated quantum codes, [arXiv:quant-ph/9608012](https://arxiv.org/abs/quant-ph/9608012).
 - [11] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing* (ACM, El Paso, TX, 1997), pp. 176–188.
 - [12] A. Y. Kitaev, Quantum computations: Algorithms and error correction, *Russ. Math. Surveys* **52**, 1191 (1997).
 - [13] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, *Science* **279**, 342 (1998).
 - [14] B. W. Reichardt, Fault-tolerance threshold for a distance-three quantum code, in *Automata, Languages and Programming*, edited by M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Lecture Notes in Computer Science, Vol. 4051 (Springer, Berlin, Heidelberg, 2006), pp. 50–61.
 - [15] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 code, *Quantum Inf. Comput.* **6**, 97 (2006).
 - [16] D. Poulin, Optimal and efficient decoding of concatenated quantum block codes, *Phys. Rev. A* **74**, 052333 (2006).
 - [17] Z. W. E. Evans and A. M. Stephens, Optimal correction of concatenated fault-tolerant quantum codes, *Quantum Info. Proc.* **11**, 1511 (2012).
 - [18] S. A. Yadavalli and I. Marvian, Noisy quantum trees: infinite protection without correction, *npj Quantum Inf.* **11**, 151 (2025).
 - [19] B. Pato, T. Tansuwanont, and K. R. Brown, Concatenated Steane code with single-flag syndrome checks, *Phys. Rev. A* **110**, 032411 (2024).

- [20] B. W. Reichardt, Improved ancilla preparation scheme increases fault-tolerant threshold, [arXiv:quant-ph/0406025](#).
- [21] A. M. Steane, Fast fault-tolerant filtering of quantum code-words, [arXiv:quant-ph/0202036](#).
- [22] A. Paetznick and B. W. Reichardt, Fault-tolerant ancilla preparation and noise threshold lower bounds for the 23-qubit Golay code, *Quantum Inf. Comput.* **12**, 1034 (2013).
- [23] A. M. Steane, Overhead and noise threshold of fault-tolerant quantum error correction, *Phys. Rev. A* **68**, 042322 (2003).
- [24] E. Knill, Quantum computing with realistically noisy devices, *Nature (London)* **434**, 39 (2005).
- [25] A. Cross, D. DiVincenzo, and B. Terhal, A comparative code study for quantum fault tolerance, *Quantum Inf. Comput.* **9**, 541 (2009).
- [26] M. Suchara, J. Kubiawicz, A. Faruque, F. T. Chong, C.-Y. Lai, and G. Paz, QuRE: The quantum resource estimator toolbox, in *2013 IEEE 31st International Conference on Computer Design (ICCD)* (IEEE, Asheville, NC, 2013), pp. 419–426.
- [27] C. Chamberland, T. Jochym-O'Connor, and R. Laflamme, Overhead analysis of universal concatenated quantum codes, *Phys. Rev. A* **95**, 022313 (2017).
- [28] T. Jochym-O'Connor and R. Laflamme, Using concatenated quantum codes for universal fault-tolerant quantum gates, *Phys. Rev. Lett.* **112**, 010505 (2014).
- [29] C. Chamberland, T. Jochym-O'Connor, and R. Laflamme, Thresholds for universal concatenated quantum codes, *Phys. Rev. Lett.* **117**, 010501 (2016).
- [30] T. J. Yoder, R. Takagi, and I. L. Chuang, Universal fault-tolerant gates on concatenated stabilizer codes, *Phys. Rev. X* **6**, 031039 (2016).
- [31] E. Nikahd, M. Sedighi, and M. Saheb Zamani, Nonuniform code concatenation for universal fault-tolerant quantum computing, *Phys. Rev. A* **96**, 032337 (2017).
- [32] C. Lin and G. Yang, Concatenated pieceable fault-tolerant scheme for universal quantum computation, *Phys. Rev. A* **102**, 052415 (2020).
- [33] H. Yamasaki and M. Koashi, Time-efficient constant-space-overhead fault-tolerant quantum computation, *Nat. Phys.* **20**, 247 (2024).
- [34] S. Yoshida, S. Tamiya, and H. Yamasaki, Concatenate codes, save qubits, *npj Quantum Inf.* **11**, 88 (2025).
- [35] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [36] D. Gottesman, Fault-tolerant quantum computation with constant overhead, *Quantum Info. Comput.* **14**, 1338 (2014).
- [37] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, High-threshold and low-overhead fault-tolerant quantum memory, *Nature (London)* **627**, 778 (2024).
- [38] G. M. Sommers, D. A. Huse, and M. J. Gullans, Dynamically generated concatenated codes and their phase diagrams, *Phys. Rev. Res.* **7**, 023086 (2025).
- [39] S. A. Moses, C. H. Baldwin, M. S. Allman, R. Ancona, L. Ascarrunz, C. Barnes, J. Bartolotta, B. Bjork, P. Blanchard, M. Bohn *et al.*, A race-track trapped-ion quantum processor, *Phys. Rev. X* **13**, 041052 (2023).
- [40] A. M. Steane, Active stabilization, quantum computation, and quantum state synthesis, *Phys. Rev. Lett.* **78**, 2252 (1997).
- [41] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
- [42] See Supplemental Material, which includes Refs. [96–101], at <http://link.aps.org/supplemental/10.1103/v477-jw81> for additional details on the spacetime code, variations on the state-preparation circuit, optimal and approximate decoding via tensor networks, implementation on Quantinuum System H2, the quantum memory test, and code switching.
- [43] C. Cao and B. Lackey, Quantum Lego: Building quantum error correction codes from tensor networks, *PRX Quantum* **3**, 020332 (2022).
- [44] C. Cao and B. Lackey, Quantum weight enumerators and tensor networks, *IEEE Trans. Inf. Theory* **70**, 3512 (2024).
- [45] C. Cao, M. J. Gullans, B. Lackey, and Z. Wang, Quantum lego expansion pack: Enumerators from tensor networks, *PRX Quantum* **5**, 030313 (2024).
- [46] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493(R) (1995).
- [47] N. H. Nguyen, M. Li, A. M. Green, C. Huerta Alderete, Y. Zhu, D. Zhu, K. R. Brown, and N. M. Linke, Demonstration of Shor encoding on a trapped-ion quantum computer, *Phys. Rev. Appl.* **16**, 024057 (2021).
- [48] The way the circuit is drawn may look familiar to practitioners of ZX calculus [94], and indeed if we absorb the stabilizer inputs into the gates, the corresponding ZX diagram is a binary tree composed of alternating layers of Z and X spiders, as discussed in Ref. [38]. Here, we leave the inputs explicit as they can also suffer errors in our noise model.
- [49] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, *Phys. Rev. A* **54**, 1098 (1996).
- [50] A. Steane, Multiple-particle interference and quantum error correction, *Proc. R. Soc. London, Ser. A* **452**, 2551 (1996).
- [51] This stabilizer will henceforth be referred to as the “root stabilizer” to distinguish it from the stabilizer inputs that enter each node on the right side. Under the encoding circuit, it evolves to the operator $\bar{P}$.
- [52] F. T. Leighton, *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes* (Morgan Kaufmann Publishers Inc., San Francisco, CA, 1992).
- [53] J. W. Cooley and J. W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Comp.* **19**, 297 (1965).
- [54] C. J. Weinstein, Quantization effects in digital filters, Technical Report, 1969.
- [55] A. J. Ferris and D. Poulin, Tensor networks and quantum error correction, *Phys. Rev. Lett.* **113**, 030501 (2014).
- [56] N. Delfosse and A. Paetznick, Spacetime codes of Clifford circuits, [arXiv:2304.05943](#).
- [57] Each measurement error occurs immediately after a two-qubit gates, so the ancilla legs exiting a gate have two error opportunities. A Pauli error immediately preceding measurement can be interpreted as a bit flip error on the classical measurement outcome.
- [58] D. Bacon, S. T. Flammia, A. W. Harrow, and J. Shi, Sparse quantum codes from quantum circuits, *IEEE Trans. Inf. Theory* **63**, 2464 (2017).
- [59] D. Gottesman, Opportunities and challenges in fault-tolerant quantum computation, [arXiv:2210.15844](#).

- [60] The precise number of spacetime qubits for this error model $(5T + \lfloor (T+1)/2 \rfloor - 2)2^{T-1}$ for $|\bar{0}\rangle$, $(5T + \lfloor (T+1)/2 \rfloor - 2)2^{T-1}$ for $|\bar{+}\rangle$. For simplicity, the scaling collapse in Fig. 2(a)(ii) drops the subleading factors and simply defines $N(T) = T2^T$. This convention does not affect the inferred exponent ν .
- [61] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, *npj Quantum Inf.* **4**, 55 (2018).
- [62] F. Venn, J. Behrends, and B. Béri, Coherent-error threshold for surface codes from Majorana delocalization, *Phys. Rev. Lett.* **131**, 060603 (2023).
- [63] K. M. R. Audenaert and M. B. Plenio, When are correlations quantum?—Verification and quantification of entanglement by simple measurements, *New J. Phys.* **8**, 266 (2006).
- [64] A. M. Stephens, Z. W. E. Evans, S. J. Devitt, and L. C. L. Hollenberg, Asymmetric quantum error correction via code conversion, *Phys. Rev. A* **77**, 062335 (2008).
- [65] A. Paetznick and B. W. Reichardt, Universal fault-tolerant quantum computation with only transversal gates and error correction, *Phys. Rev. Lett.* **111**, 090505 (2013).
- [66] J. T. Anderson, G. Duclos-Cianci, and D. Poulin, Fault-tolerant conversion between the Steane and Reed-Muller quantum codes, *Phys. Rev. Lett.* **113**, 080501 (2014).
- [67] A. Kubica and M. E. Beverland, Universal transversal gates with color codes: A simplified approach, *Phys. Rev. A* **91**, 032330 (2015).
- [68] Héctor Bombín, Dimensional jump in quantum error correction, *New J. Phys.* **18**, 043038 (2016).
- [69] C. Vuillot, L. Lao, B. Criger, C. García Almudéver, K. Bertels, and B. M. Terhal, Code deformation and lattice surgery are gauge fixing, *New J. Phys.* **21**, 033028 (2019).
- [70] I. Pogorelov, F. Butt, L. Postler, C. D. Marciniak, P. Schindler, M. Müller, and T. Monz, Experimental fault-tolerant code switching, *Nat. Phys.* **21**, 298 (2025).
- [71] D. Gottesman, Surviving as a quantum computer in a classical world, Draft (2024), <https://www.cs.umd.edu/class/spring2024/cmssc858G/>.
- [72] In some works, including Ref. [41], the state we call $|T\rangle$ is instead called the “Hadamard magic state” $|H\rangle$, while $|T\rangle$ is reserved for an eigenstate of the T gate. Elsewhere, e.g., in Ref. [35], it is denoted $|A\rangle$. We use $|T\rangle$ (as in, e.g. Ref. [75]), to emphasize that it is a resource state for the T gate.
- [73] A protocol subsequently developed in Ref. [95] achieves an optimal distillation threshold of $(1 - 1/\sqrt{2})/2 = 0.85355, \dots$, but has inferior subthreshold performance to the scheme of Ref. [41]. We propose using the distillation scheme of Ref. [41] since the fixed-point logical fidelity on the singletree already exceeds its threshold in the entire coding phase.
- [74] Eagle-eyed readers will note that this figure uses the opposite encoding order (GS_e rather than GS_o) from the rest of the text. We made this choice due to an odd-even effect in the fixed-point logical fidelity, which results in inferior performance of GS_o for this basis of logical T state at even depths. See Sec. III C of the Supplemental Material [42] for further details.
- [75] C. Gidney, N. Shutty, and C. Jones, Magic state cultivation: Growing T states as cheap as CNOT gates, [arXiv:2409.17595](https://arxiv.org/abs/2409.17595).
- [76] K. M. Svore, B. M. Terhal, and D. P. DiVincenzo, Local fault-tolerant quantum computation, *Phys. Rev. A* **72**, 022317 (2005).
- [77] K. Svore, A. Cross, I. Chuang, and A. Aho, A flow-map model for analyzing pseudothresholds in fault-tolerant quantum computing, *Quantum Inf. Comput.* **6**, 193 (2006).
- [78] Y. Wu, S. Kolkowitz, S. Puri, and J. D. Thompson, Erasure conversion for fault-tolerant quantum computing in alkaline earth Rydberg atom arrays, *Nat. Commun.* **13**, 4657 (2022).
- [79] K. Sahay, J. Jin, J. Claes, J. D. Thompson, and S. Puri, High-threshold codes for neutral-atom qubits with biased erasure errors, *Phys. Rev. X* **13**, 041013 (2023).
- [80] M. Kang, W. C. Campbell, and K. R. Brown, Quantum error correction with metastable states of trapped ions using erasure conversion, *PRX Quantum* **4**, 020358 (2023).
- [81] A. Kubica, A. Haim, Y. Vaknin, H. Levine, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the T_1 limit in superconducting circuits, *Phys. Rev. X* **13**, 041022 (2023).
- [82] J. D. Teoh, P. Winkel, H. K. Babla, B. J. Chapman, J. Claes, S. J. de Graaf, J. W. O. Garmon, W. D. Kalfus, Y. Lu, A. Maiti, K. Sahay, N. Thakur, T. Tsunoda, S. H. Xue, L. Frunzio, S. M. Girvin, S. Puri, and R. J. Schoelkopf, Dual-rail encoding with superconducting cavities, *Proc. Natl. Acad. Sci. USA* **120**, e2221736120 (2023).
- [83] P. Scholl, A. L. Shaw, R. B.-S. Tsai, R. Finkelstein, J. Choi, and M. Endres, Erasure conversion in a high-fidelity Rydberg quantum simulator, *Nature (London)* **622**, 273 (2023).
- [84] S. Ma, G. Liu, P. Peng, B. Zhang, S. Jandura, J. Claes, A. P. Burgers, G. Pupillo, S. Puri, and J. D. Thompson, High-fidelity gates and mid-circuit erasure conversion in an atomic qubit, *Nature (London)* **622**, 279 (2023).
- [85] H. Levine, A. Haim, J. S. C. Hung, N. Alidoust, M. Kalae, L. DeLorenzo, E. A. Wollack, P. Arrangoiz-Arriola, A. Khalajhedayati, R. Sanil *et al.*, Demonstrating a long-coherence dual-rail erasure qubit using tunable transmons, *Phys. Rev. X* **14**, 011051 (2024).
- [86] K. S. Chou, T. Shemma, H. McCarrick, T.-C. Chien, J. D. Teoh, P. Winkel, A. Anderson, J. Chen, J. C. Curtis, S. J. de Graaf *et al.*, A superconducting dual-rail cavity qubit with erasure-detected logical measurements, *Nat. Phys.* **20**, 1454 (2024).
- [87] A. Koottandavida, I. Tsioutsios, A. Kargioti, C. R. Smith, V. R. Joshi, W. Dai, J. D. Teoh, J. C. Curtis, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret, Erasure detection of a dual-rail qubit encoded in a double-post superconducting cavity, *Phys. Rev. Lett.* **132**, 180601 (2024).
- [88] C. M. Holland, Y. Lu, S. J. Li, C. L. Welsh, and L. W. Cheuk, Demonstration of measurement-enhanced state preparation and erasure conversion in a molecular tweezer array, *Phys. Rev. X* **15**, 031018 (2025).
- [89] B. Zhang, G. Liu, G. Bornet, S. P. Horvath, P. Peng, S. Ma, S. Huang, S. Puri, and J. D. Thompson, Leveraging erasure errors in logical qubits with metastable ^{171}Yb atoms, [arXiv:2506.13724](https://arxiv.org/abs/2506.13724).
- [90] R. Singleton, An algorithm for computing the mixed radix fast Fourier transform, *IEEE Trans. Audio Electroacoust.* **17**, 93 (1969).

- [91] S. Krastanov, F. A. Mian, Z. Ni, P. Viswanathan, S. Pardis, A. Micciche, J. Lapeyre, Y. Yan, T. Hofmann, gsommers *et al.*, QuantumSavory/QuantumClifford.jl, Zenodo (2025), <https://doi.org/10.5281/zenodo.5208167>.
- [92] <https://github.com/gsommers/multitree/>.
- [93] Note that without the classical register, $\mathcal{E}_{X,e}$ would be equivalent to $\mathcal{N}_{X,e/2}$, while for $\theta = \pi/2$, the Pauli twirl of $\mathcal{C}_{X,q,\pi/2}$ would match $\mathcal{E}_{X,q}$.
- [94] J. van de Wetering, ZX-calculus for the working quantum computer scientist, [arXiv:2012.13966](https://arxiv.org/abs/2012.13966).
- [95] B. W. Reichardt, Quantum universality from magic states distillation applied to CSS codes, *Quantum Info. Proc.* **4**, 251 (2005).
- [96] D. Poulin, Stabilizer formalism for operator quantum error correction, *Phys. Rev. Lett.* **95**, 230504 (2005).
- [97] V. V. Albert, Boulder lecture notes: Modern quantum error correction with 4 qubits, 2023, <https://www.colorado.edu/conference/bss/media/553>.
- [98] M. J. Gullans, S. Krastanov, D. A. Huse, L. Jiang, and S. T. Flammia, Quantum coding with low-depth random circuits, *Phys. Rev. X* **11**, 031066 (2021).
- [99] A. Kulkarni and B. Lackey, Quantum circuit tensors and enumerators with applications to quantum fault tolerance, *IEEE Trans. Inf. Theory* **71**, 4406 (2025).
- [100] OEIS Foundation Inc., Entry A000695: Central trinomial coefficients, The On-line Encyclopedia of Integer sequences, 2025, <https://oeis.org/A000695>.
- [101] D. Bacon, Operator quantum error-correcting subsystems for self-correcting quantum memories, *Phys. Rev. A* **73**, 012340 (2006).