Supplemental Material: Sharing flag circuits across code blocks: an exact local criterion for joint logical measurements

Annan Chen

This supplement gives finite circuit certificates, details of the five-qubit-code experiment and the complete Steane protocol, followed by Shor-code noise calculations, verification methods and comparisons with earlier constructions. The code and data behind every table and figure are archived [1]. Cross-references to the main text retain its labels; supplementary sections, tables, figures and equations are prefixed S.

S1 Weight-five obstruction and the four-flag lower bound

S1.1 A two-flag example

Consider the gate sequence

$$F_2\ D_1\ D_2\ F_1\ D_3\ D_4\ F_2\ D_5\ F_1.$$

Prepare the measurement qubit in $|+\rangle$, the flags in $|0\rangle$, and measure them in X and Z, respectively. Even flag couplings give G0 by Lemma 1. With records printed as $f_1 f_2$, the corrections admitted by the X-weight guarantee X_2 are $00 : \{I\}$, $10 : \{I, X_5\}$, $01 : \{X_1\}$ and $11 : \{X_1X_2X_3\}$, modulo M. Choose $00 \mapsto I$, $10 \mapsto I$, $01 \mapsto X_1$, and $11 \mapsto X_1X_2X_3$. This circuit satisfies X_2 and L_1, but admits no L_2 correction. Its relation to Fig. 8 and Table VI of Anker–Marvian [2] is documented in Section S7.1.

Order layer 1 as A, B and layer 3 as B, A. Place an X fault on the measurement qubit after the gate to $q_{A,1}$ and another after the gate to $q_{B,3}$. The flags are $(1, 0)$, so the displayed correction is I, leaving X_2 on A and $X_1X_2X_3$ on B. The two common choices give block weights $(1, 3)$ and $(4, 2)$. In the Shor $[[25, 1, 5]]$ code (Section S5), take the support to be one five-qubit repetition row: $X_1X_2X_3$ and X_4X_5 have the same ZZ-check syndrome, so correcting the latter leaves a logical X on B. For this sequence and the second even-coupling sequence in Section S7.1, all sixteen two-block orders modulo exchange fail even with different block corrections. Section S6.4 describes the physical events that exclude every correction pair for these layer orders.

Figure S1 shows this two-fault history.

S1.2 A stronger weight-five lower bound

The classified family fixes one measurement qubit and the data order $D_1, \ldots, D_5$, each coordinate occurring once. A data boundary is a slot between successive D_j gates or at either end of the sequence. Each flag couples exactly twice at distinct data boundaries; all orders of different flags within a boundary are included. Flag labels are normalized by their first appearance. Deterministic recovery accepts every record reachable within the fault budget.

Proposition S1 (X-weight lower bound at fixed layer orders)**.** *At $w = 5, t = 2, m = 2$, no circuit in this family with $k \leq 3$ flags has a deterministic X-weight correction certificate. This holds even if each circuit chooses its own fixed layer order and each record chooses arbitrary separate X corrections $r_A(f), r_B(f)$. Four flags suffice for the stronger Pauli-weight guarantee and ideal G0, for every complete lift.*

With acceptance restricted to $f = 0$, the two-flag circuit $D_1F_1D_2F_2D_3D_4F_1D_5F_2$ satisfies L_2 with the identity correction.

Computer-assisted proof. Two independent enumerators give the coverage in Table S1. For each sequence, global exchange of A and B fixes the first layer and leaves sixteen orders. One checker intersects all $32^2 = 1024$ correction pairs at each record; the other uses 512 representatives modulo simultaneous complementation. Every case fails. The four-flag circuit of Fig. 4 has a complete physical L_2 check and satisfies Lemma 1. Section S6.4 proves the enumeration and fault-alphabet coverage and explains the physical witness checks. □

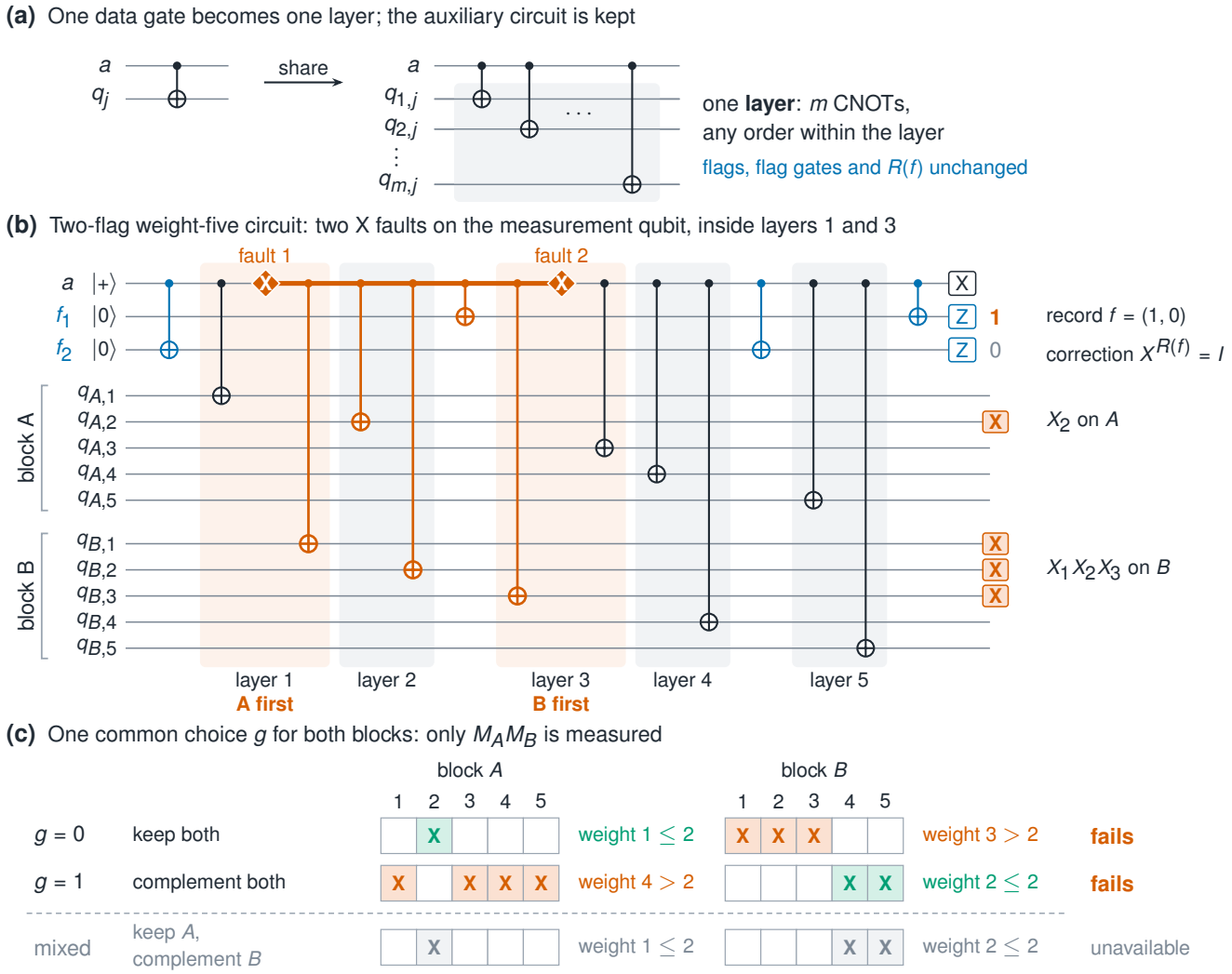


Figure S1: A two-fault history in the weight-five circuit. The gate order, fault positions, record and output residuals are those of the text. The table lists the two common choices and the unavailable mixed one.

Table S1: Exhaustive deterministic X-weight correction classification for two blocks at $w = 5$, $t = 2$. Each sequence has sixteen layer orders modulo block exchange; all reachable records must be accepted.

Flags	Canonical sequences	Fixed-order cases	Feasible
0	1	16	0
1	15	240	0
2	195	3,120	0
3	2,605	41,680	0
Total	2,816	45,056	0

On a single block, with every reachable record accepted, the minimum is two flags for X weight and four for Pauli weight at $d = 5$, and zero and one at $d = 3$. Section S6 classifies larger families; at distance seven the Pauli-weight minimum lies between six and nine flags.

S2 Interval constructions and further certificates

S2.1 Interval circuits and their verified range

For every interval $[i, i+L]$ with $L = 2, \ldots, t$ and $i = 0, \ldots, w-L$, attach one flag at its endpoints. At a common boundary use either closings before openings or openings before closings, with a fixed increasing-length order within each group. At $w = 2t+1$ this uses

$$k = \frac{(t-1)(3t+2)}{2}, \qquad N_{\mathrm{CNOT}} = wm + 2k.$$

The $t = 2$ member is the four-flag circuit. The $t = 1$ sequence $F_1 D_1 D_2 F_1 D_3$ is treated separately. Table S2 gives the certificates and their completed orders.

Table S2: Pauli-weight correction certificates. Each completed order has an exhaustive effect check. Flag counts give resource upper bounds at the completed order. The final row is complete through order four.

Target t	w	Flags	Completed order	Boundary orders
2	5	4	2	both
3	7	11	3	both
3	7	9	3	listed sequence
4	9	21	4	both
4	9	14	4	listed sequence
5	11	34	5	both
6	13	50	4	closings before openings

All 2,047 nonempty subsets of the eleven distance-seven windows were checked in both boundary orders. Passing sequences with nine, ten, and eleven flags number eight, twelve, and two. Nine is therefore the minimum within that subfamily. For all circuits with one measurement qubit and twice-coupled flags, the exhaustive five-flag exclusion at Pauli-weight order two only supplies the lower bound six at order three. Four deterministic greedy deletion orders at distance nine yielded a fourteen-flag subfamily; its complete fixed-table four-fault check covers 3,764,376 order-four effect multisets.

Conjecture S1 (Interval family)**.** *For every $t \geq 2$, the two boundary-order constructions above at $w = 2t+1$ admit a Pauli-weight correction certificate through order t.*

Complete verification covers $t = 2, 3, 4, 5$. For $t = 6$, fifty flags were checked completely only through order four in the closing-before-opening order.

At a data cut j, an interval row has value $h_{i,L}(j) = 1$ exactly when $i < j \leq i + L$; its discrete derivative has the two endpoints as support. This parity structure gives the interval rows of Anker–Marvian's classical-code framework [2]; the verified circuits provide resource upper bounds.

When a suitable heavier logical representative exists, an alternative is to combine an existing X-weight certificate at $w \geq d+1$ with main-text Proposition 8. At $d = 5$ the $w = 5$ and $w = 6$ routes use four and five flags, and $5m + 8$ and $6m + 10$ CNOTs, respectively.

S2.2 The two listed sequences

The rows of Table S2 marked "listed sequence" use nine flags at $w = 7$ and fourteen flags at $w = 9$:

$$\begin{aligned} w = 7,\ 9 \text{ flags}: \quad & F_1F_2D_1\ \ F_3F_4D_2\ \ F_1F_5F_6D_3\ \ F_3F_2F_7D_4 \\ & F_5F_4F_8D_5\ \ F_7F_6F_9D_6\ \ F_8D_7\ \ F_9, \\ w = 9,\ 14 \text{ flags}: \quad & F_1D_1\ \ F_2F_3D_2\ \ F_4F_5D_3\ \ F_2F_6F_7D_4 \\ & F_4F_1F_8F_9F_{10}D_5\ \ F_6F_3F_{11}D_6\ \ F_8F_5F_{12}F_{13}D_7 \\ & F_9F_7F_{14}D_8\ \ F_{12}F_{10}D_9\ \ F_{14}F_{13}F_{11}. \end{aligned}$$

S3 Implementation of the five-qubit-code experiment

The single-round primitive analyzed in Section VII is the single-round circuit in Freund *et al.*'s Fig. 12 [3]. Their public function `ft_measurement_joint_logical_X` bundles two measurement rounds with one flag initialized

once and measured only at the end [4]. It contains 16 native Mølmer–Sørensen (MS) gates, 26 one-qubit gates and three final Z readouts. The controller accepts agreement of the two measurement outcomes and an untriggered aggregate flag. A separately read flag after each round gives the single-round recovery rule of Section VII; the experimental bundle uses the combined two-round record.

Appendix D and Fig. 13 of Ref. [3] distinguish deterministic error correction from error detection with rejection. The reported teleportation experiment uses the latter. Appendix G gives the experiment's syndrome-conditioned recovery table. To use the local correction $\mathcal{R}(1) = Z_1$ in a deterministic controller, the subsequent syndrome and correction table must be transformed consistently: for a stabilizer generator G_i, its recorded bit changes by the commutation bit of G_i and $\mathcal{R}(f)$, and the net correction includes that correction.

S4 Complete Steane protocol and sampling

S4.1 Code, records and physical controller

The positive stabilizers, in coordinate order $1, \ldots, 7$, are

$$IIIXXXX, \quad IXXIIXX, \quad XIXIXIX, \qquad IIIZZZZ, \quad IZZIIZZ, \quad ZIZIZIZ.$$

We use $X_L = X_1X_2X_3$ and $Z_L = Z_1Z_2Z_3$. The six checks are scanned in the displayed order. For a weight-four X check with ordered support (j_1, j_2, j_3, j_4), prepare the measurement qubit in $|+\rangle$ and the flag in $|0\rangle$, and apply $D_{j_1}FD_{j_2}D_{j_3}FD_{j_4}$, with CNOTs controlled by the measurement qubit. Read the measurement qubit in X and flag in Z. For a Z check reverse every CNOT, prepare the measurement qubit in $|0\rangle$ and flag in $|+\rangle$, and exchange the readout bases. This is the CSS counterpart of Chao–Reichardt Fig. 3(b) [5]. Basis preparation and readout are implemented by a reset to $|0\rangle$, H where required, and Z readout; each of these operations has its own noise.

At the first nonzero interpreted syndrome or flag, perform unflagged measurements of all six checks. For no raised flag, use the unique single-coordinate correction for each of the X and Z syndrome components. For a raised flag, use the table for the triggering check. This table is obtained by enumerating every single physical fault in that check, propagating the data residual, and grouping by the six repeated syndromes. Within each group all residuals differ only by code stabilizers. All remaining entries use ordinary CSS correction. The six tables are archived; the stabilizer bit order is the order above. The standalone QEC check covers 4,068 single-fault assignments and all 21 single-qubit input Paulis.

Each quantum operation, classical decision and Pauli-frame update takes one time step (tick). Pauli-frame updates are classical operations during which data and auxiliaries idle. The same pair of auxiliaries is reused for joint measurements and the serial QEC modules. Both remain allocated for all ticks, including before preparation and after measurement; resets remove their previous errors. A QEC scan that stops after check $j+1$, $0 \le j \le 5$, takes $56 + 13j$ ticks including its repeated checks and recovery decision. A clean six-check scan takes 78 ticks. The longest complete paths have 759 ticks for bare and 771 for flagged measurement. The executed path is used for both noise and resources. The controller graph, the gate lists of all conditional branches and the individual measurement templates are archived with the executable model.

Figure S2(a) shows the three rounds and their interpreted outcomes. Quantum blocks continue through the QEC modules; the three classical votes determine the reported bit.

S4.2 Instrument and physical-output criterion

Write E for the effective data Pauli after tracked corrections. Since every ideal data measurement is either a code stabilizer or M_AM_B, each physical Pauli trajectory has a Kraus operator proportional to $E\Pi_y$ on the input codespace. The recorded readout flips and ideal branch y are tracked separately. An ideal CSS decoder, used only to assess the completed output, removes the final syndrome. The remaining logical Pauli has four binary components (x_A, z_A, x_B, z_B). Its action on $\Pi_y\rho\Pi_y$ is trivial for arbitrary input and reference precisely when $z_A = z_B = 0$ and $x_A = x_B$. The identity class is generated by the single measured product. The classical report must also equal y.

The physical one-fault certificate additionally checks

$$\min_{g\in\{0,1\}} \max_{b\in\{A,B\}} \min_{G\in\mathcal{G}_b} \mathrm{wt}(GE_bM_b^g) \le 1,$$

where $\mathcal{G}_b$ is the stabilizer group of block b. For the single-input-error, fault-free cases, the bound is zero. The flagged implementation also has zero residual on all 441 pairs of input Paulis, one in each block, when there are

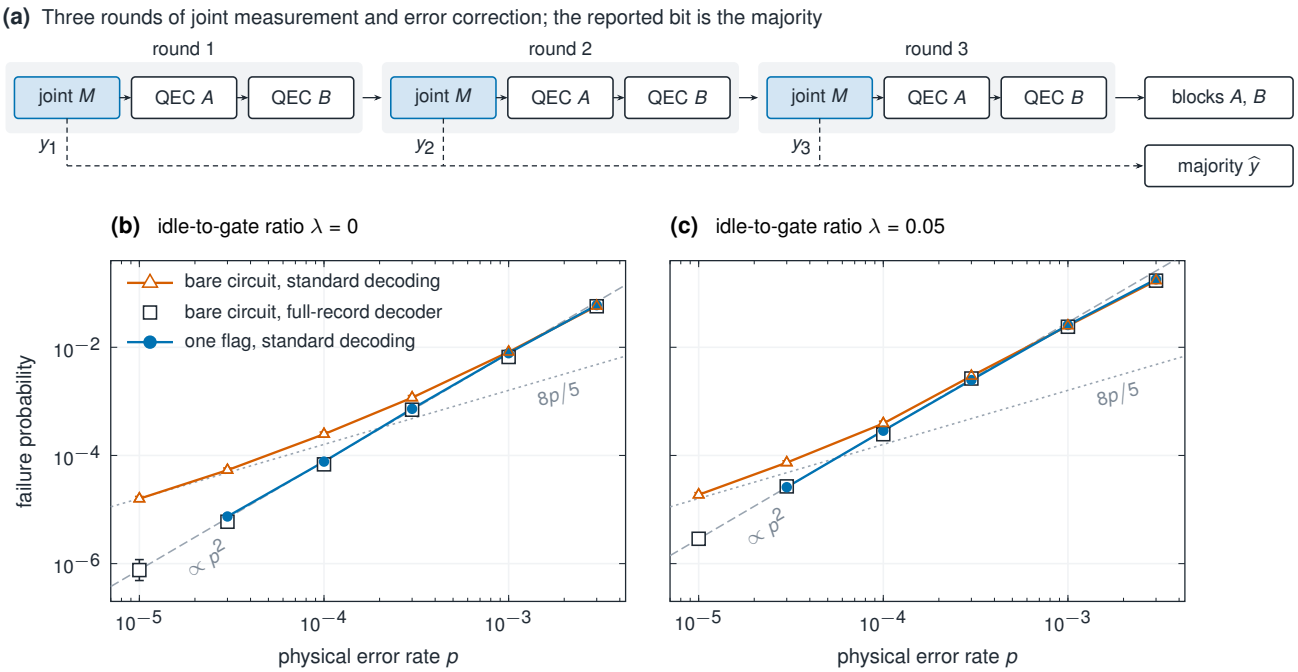


Figure S2: Two-block Steane protocol and all sampled failure probabilities. (a) Each round applies the joint measurement and then error correction on each block; the reported bit is the majority of the interpreted outcomes y_1, y_2, y_3. (b),(c) The 34 plotted estimates, with counts in Tables S4 and S5, for idle-to-gate error ratios $\lambda = 0$ and 0.05, with the first-order term $8p/5$ (dotted) and a quadratic guide through the one-flag estimate at $p = 10^{-4}$ (dashed). Bars are 95% Wilson intervals, most of them smaller than the symbols. Panel (b) is main-text Fig. 5(a).

no internal faults. This larger input set is recorded separately from the common $u + s \leq 1$ comparison, where u is the total input Pauli weight and s the internal fault count.

All full-Pauli faults at resets, readouts, H gates, CNOTs and idles are enumerated. At reset and readout some Pauli assignments are operationally equivalent or harmless; they remain in the certificate count. The probabilistic model uses the effective preparation/readout flip of total probability p. For a clean input and exactly one internal fault, the prefix before that fault follows the no-fault control path. Enumerating every position on that path and then following the fault-dependent continuation therefore covers every permitted single-fault trajectory, including its adaptive QEC branch. Fault-free erroneous inputs are enumerated separately. Both values of the ideal joint outcome are included when forming the record decoder.

For the bare protocol, the full record determines a compatible pair of reported bit and logical correction on all 2,336 record classes reached with at most one input error or one internal fault. This table is the full-record decoder of the bare protocol. Each row contains the reported bit and three correction bits selecting $X_{L,A}, Z_{L,A}, Z_{L,B}$, and a record outside the table keeps the reported bit and correction of standard decoding. The analogous table of the flagged protocol has 2,396 classes; on each of them it gives the reported bit of standard decoding and no additional logical correction, so for the flagged protocol the full-record decoder coincides with standard decoding at all fault orders. An independent implementation replays all 53,024 events with $u + s \leq 1$ and checks the record, output class, reported bit, duration and physical residual bound. Independent encoded-Choi simulations with Stim [6] replay twenty selected physical trajectories for each ideal joint outcome, including flag triggers, repeated QEC, inputs and the bare ambiguity.

Table S3: Coverage of the physical Steane certificates and independent checks. Single-fault counts include all admitted Pauli assignments at each physical location. Input tests have no internal faults.

Check	Bare	Flagged
Single-fault assignments	26,154	26,784
Single-input Pauli errors	42	42
Input pairs, one Pauli per block	—	441
Full-record decoder classes	2,336	2,396
Standard-decoder single-fault failures	24	0
Standalone QEC: faults / input errors	4,068 / 21	
Independent replay, trajectories	53,024	
Stim encoded-Choi branches	40	

S4.3 Sampling and resources

At every H gate the post-gate error is uniform on X, Y, Z with total probability p. At a CNOT it is uniform on the fifteen nonidentity two-qubit Paulis with total probability p. Resets have an X flip after preparation; Z readouts have a bit flip before readout, each with probability p. Every idle has uniform one-qubit depolarization of total probability λp. Preparation of the arbitrary encoded input lies outside this measurement protocol's cost boundary. Classical arithmetic and software correction storage are reliable.

The sampler draws independent Bernoulli faults along the executed adaptive path. Geometric skipping of fault-free visited locations implements this distribution, separately for operation and idle probabilities. A direct implementation drawing one Bernoulli variate per visited location was also run for 100,000 trajectories per protocol at $p = 10^{-3}, \lambda = 0.05$; it gave 2,427 and 2,586 failures, consistent with the geometric implementation's 2,398 and 2,581.

For the sixteen-cell base-range dataset, independent 20,000-shot pilots set the sample sizes. It contains 9,380,000 trajectories, with a fixed budget of 100,000 to 3,000,000 per cell (Table S4). Each point has a fixed-N Wilson 95% interval. The archived data also give Bonferroni simultaneous 95% intervals for the sixteen points and Newcombe confidence intervals for differences between independent protocol estimates at matching parameters. Counts for wrong reports and wrong logical outputs are given separately; their union defines failure. All trajectories complete and are retained. Sampling fixes the underlying ideal outcome to $y = 0$: changing it to one complements all three raw joint bits, leaves every QEC record and branch unchanged, and complements the decoder's reported bit. The full-record table contains these paired records with identical logical corrections, and the majority vote used outside the table has the same symmetry. The failure criterion therefore has the same probability in either ideal branch, including for logical inputs entangled with a reference. Both branches were re-executed for all 53,024 certificate cases. A further 20,000 paired noisy trajectories at $p = 3 \times 10^{-3}$, $\lambda = 0.05$, including 12,737 pairs with records outside the table, obey the same covariance.

Resource totals, mean depth and depth quantiles are accumulated from the executed branches; Table S7 lists the resources of error-free (clean) runs and the means at $p = 10^{-3}$. Both protocols use 16 qubits, so the clean depths of 501 and 513 ticks correspond to 8,016 and 8,208 qubit-ticks. At $p = 10^{-4}$, $\lambda = 0$, for example, the mean depths are 501.444 and 513.513 ticks. The archived data give the finite-noise means and the 50th and 95th percentiles for every cell.

For the three-decoder comparison, twelve bare cells use $p = 10^{-5}, 3 \times 10^{-5}, 10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}$ at both idle ratios; two flagged cells use $p = 3 \times 10^{-5}$. Budgets were fixed before these runs to target 400 expected failures (about a 9.8% relative 95% half-width), rounded up to 100,000 shots, with a minimum of 100,000. The planning rates use the exact $8p/5$ term at the two lowest bare points and independent base-range counts elsewhere; the flagged low-noise forecast scales the base-range $p = 10^{-4}$ rate quadratically. These fourteen runs contain 143,900,000 trajectories. Each primary estimate (standard bare or flagged) has at least 390 failures and a relative 95% interval half-width below 10%. The sampling plan is archived. Both bare decoders are evaluated on each new trajectory; those estimates are paired. Figure S2 combines the sixteen base-range estimates with twelve standard bare estimates, four low-p full-record bare estimates and two low-p flagged estimates. At the two lowest bare rates, both decoder estimates share trajectories; the remaining plotted estimates are independent. Table S5 gives all counts of the three-decoder dataset, including paired full-record estimates at rates also sampled by the base-range dataset.

At $p = 3 \times 10^{-5}$ and $\lambda = 0$, standard bare decoding gives 449 failures in 8,400,000 trajectories and flagged decoding gives 430 in 57,800,000. Their estimated risk ratio is 7.185. For these independent binomial sam-

Table S4: Base-range dataset: fixed-sample Monte Carlo for the bare protocol with its full-record decoder and for the flagged protocol. k of N shots fail; intervals are pointwise Wilson 95%.

Protocol	λ	p	N	k	P_{fail} [95% interval]
Bare, full-record decoder	0	10^{-4}	2,000,000	137	$(6.85\ [5.80, 8.10]) \times 10^{-5}$
Flagged	0	10^{-4}	3,000,000	231	$(7.70\ [6.77, 8.76]) \times 10^{-5}$
Bare, full-record decoder	0	3×10^{-4}	750,000	526	$(7.01\ [6.44, 7.64]) \times 10^{-4}$
Flagged	0	3×10^{-4}	600,000	437	$(7.28\ [6.63, 8.00]) \times 10^{-4}$
Bare, full-record decoder	0	10^{-3}	100,000	662	$(6.62\ [6.14, 7.14]) \times 10^{-3}$
Flagged	0	10^{-3}	100,000	777	$(7.77\ [7.24, 8.33]) \times 10^{-3}$
Bare, full-record decoder	0	3×10^{-3}	100,000	5,738	$(5.74\ [5.60, 5.88]) \times 10^{-2}$
Flagged	0	3×10^{-3}	100,000	6,024	$(6.02\ [5.88, 6.17]) \times 10^{-2}$
Bare, full-record decoder	0.05	10^{-4}	1,000,000	247	$(2.47\ [2.18, 2.80]) \times 10^{-4}$
Flagged	0.05	10^{-4}	1,000,000	287	$(2.87\ [2.56, 3.22]) \times 10^{-4}$
Bare, full-record decoder	0.05	3×10^{-4}	130,000	344	$(2.65\ [2.38, 2.94]) \times 10^{-3}$
Flagged	0.05	3×10^{-4}	100,000	248	$(2.48\ [2.19, 2.81]) \times 10^{-3}$
Bare, full-record decoder	0.05	10^{-3}	100,000	2,398	$(2.40\ [2.30, 2.49]) \times 10^{-2}$
Flagged	0.05	10^{-3}	100,000	2,581	$(2.58\ [2.48, 2.68]) \times 10^{-2}$
Bare, full-record decoder	0.05	3×10^{-3}	100,000	17,071	$(1.71\ [1.68, 1.73]) \times 10^{-1}$
Flagged	0.05	3×10^{-3}	100,000	18,085	$(1.81\ [1.78, 1.83]) \times 10^{-1}$

Table S5: Three-decoder dataset. For the bare protocol both decoders are evaluated on the same N trajectories; k is the number of failures and intervals are pointwise Wilson 95%.

			Standard decoding		Full-record decoder	
λ	p	N	k	P_{fail} [95% interval]	k	P_{fail} [95% interval]
Bare protocol						
0	10^{-5}	25,000,000	396	$(1.58\ [1.44, 1.75]) \times 10^{-5}$	19	$(7.60\ [4.87, 11.87]) \times 10^{-7}$
0	3×10^{-5}	8,400,000	449	$(5.35\ [4.87, 5.86]) \times 10^{-5}$	50	$(5.95\ [4.52, 7.85]) \times 10^{-6}$
0	10^{-4}	1,800,000	445	$(2.47\ [2.25, 2.71]) \times 10^{-4}$	159	$(8.83\ [7.56, 10.32]) \times 10^{-5}$
0	3×10^{-4}	400,000	468	$(1.17\ [1.07, 1.28]) \times 10^{-3}$	299	$(7.47\ [6.67, 8.37]) \times 10^{-4}$
0	10^{-3}	100,000	819	$(8.19\ [7.65, 8.77]) \times 10^{-3}$	705	$(7.05\ [6.55, 7.59]) \times 10^{-3}$
0	3×10^{-3}	100,000	5,782	$(5.78\ [5.64, 5.93]) \times 10^{-2}$	5,672	$(5.67\ [5.53, 5.82]) \times 10^{-2}$
0.05	10^{-5}	25,000,000	469	$(1.88\ [1.71, 2.05]) \times 10^{-5}$	72	$(2.88\ [2.29, 3.63]) \times 10^{-6}$
0.05	3×10^{-5}	8,400,000	619	$(7.37\ [6.81, 7.97]) \times 10^{-5}$	225	$(2.68\ [2.35, 3.05]) \times 10^{-5}$
0.05	10^{-4}	1,000,000	390	$(3.90\ [3.53, 4.31]) \times 10^{-4}$	260	$(2.60\ [2.30, 2.94]) \times 10^{-4}$
0.05	3×10^{-4}	200,000	587	$(2.94\ [2.71, 3.18]) \times 10^{-3}$	507	$(2.53\ [2.32, 2.77]) \times 10^{-3}$
0.05	10^{-3}	100,000	2,494	$(2.49\ [2.40, 2.59]) \times 10^{-2}$	2,422	$(2.42\ [2.33, 2.52]) \times 10^{-2}$
0.05	3×10^{-3}	100,000	17,064	$(1.71\ [1.68, 1.73]) \times 10^{-1}$	17,031	$(1.70\ [1.68, 1.73]) \times 10^{-1}$
Flagged protocol						
0	3×10^{-5}	57,800,000	430	$(7.44\ [6.77, 8.18]) \times 10^{-6}$		
0.05	3×10^{-5}	15,500,000	401	$(2.59\ [2.35, 2.85]) \times 10^{-5}$		

ples, the approximate log-risk-ratio (Katz) 95% interval is $[6.295, 8.201]$, obtained by exponentiating $\log(\widehat{p}_1/\widehat{p}_2) \pm 1.96\sqrt{1/k_1 - 1/N_1 + 1/k_2 - 1/N_2}$. Table S6 lists the ratios and intervals plotted in main-text Fig. 5(b).

The 24 bare-standard witnesses were replayed for both ideal outcomes. They occur at the CNOTs to $q_{A,2}$ and $q_{B,2}$ in each round. After the CNOT to $q_{A,2}$ the failing assignments are X_aI, X_aZ, Y_aI, Y_aZ, where the first factor acts on the measurement qubit a and the second on the data qubit; after the CNOT to $q_{B,2}$ they are X_aX, X_aY, Y_aX, Y_aY. Each has coefficient $1/15$ in the depolarizing model, giving $8/5$ in total. All have an incorrect logical output and a correct report. The archived location table lists their ticks and Pauli factors.

Table S6: Failure probabilities of the bare protocol relative to the one-flag protocol, computed from the counts of Tables S4 and S5 for independent samples, with 95% log-ratio intervals. The standard-decoding column is plotted in main-text Fig. 5(b).

λ	p	Bare, standard decoding	Bare, full-record decoder
0	3×10^{-5}	7.18 [6.29, 8.20]	0.80 [0.60, 1.07]
0	10^{-4}	3.21 [2.74, 3.76]	0.89 [0.72, 1.10]
0	3×10^{-4}	1.61 [1.41, 1.83]	0.96 [0.85, 1.09]
0	10^{-3}	1.05 [0.96, 1.16]	0.85 [0.77, 0.94]
0	3×10^{-3}	0.96 [0.93, 0.99]	0.95 [0.92, 0.99]
0.05	3×10^{-5}	2.85 [2.51, 3.23]	1.04 [0.88, 1.22]
0.05	10^{-4}	1.36 [1.17, 1.58]	0.86 [0.73, 1.02]
0.05	3×10^{-4}	1.18 [1.02, 1.37]	1.07 [0.91, 1.26]
0.05	10^{-3}	0.97 [0.92, 1.02]	0.93 [0.88, 0.98]
0.05	3×10^{-3}	0.94 [0.93, 0.96]	0.94 [0.93, 0.96]

Table S7: Steane protocol resources. Both circuits allocate 14 data qubits and two auxiliaries. Clean paths and mean executed paths at $p = 10^{-3}$ are shown. Feedback includes decision ticks; every allocated unused qubit idles. Every shot runs to completion.

Circuit	Path	Resets	H	CNOT	Readouts	Feedback	Ticks	Idles
Bare	clean	75	78	234	75	39	501	7320
Bare	mean, $\lambda = 0$	75.39	78.39	237.59	75.39	38.38	505.14	7377.87
Bare	mean, $\lambda = 0.05$	75.34	78.34	239.13	75.34	37.80	505.95	7387.86
Flagged	clean	78	78	240	78	39	513	7494
Flagged	mean, $\lambda = 0$	78.47	78.47	243.95	78.47	38.39	517.75	7560.76
Flagged	mean, $\lambda = 0.05$	78.43	78.43	245.61	78.43	37.78	518.69	7572.52

The full-record bare and flagged decoders have zero first-order coefficients by their exhaustive one-fault certificates and the bounded controller.

S5 Shor-code noise comparison

The Shor $[[25, 1, 5]]$ code has 25 data qubits per block, $q_{i,j}$ in row i and column j, with stabilizers $Z_{i,j}Z_{i,j+1}$ and $\prod_{j=1}^{5} X_{i,j}X_{i+1,j}$. The five disjoint row operators $\prod_j X_{i,j}$ therefore represent the same logical X. The comparison measures all five rows, each jointly on every block, and combines the five outcomes by majority vote. Three circuits are compared for the measurement of one row: the four-flag circuit of main-text Fig. 4(b); a cat-state circuit verified before the data interactions, after Shor [7] (*verified cat*); and a cat-state circuit decoded after the data interactions, after DiVincenzo and Aliferis [8] (*decoded cat*).

S5.1 Time schedules

Figure S3 shows the cat preparation and the two schedules. Every operation takes one tick, and operations in a tick have disjoint supports. Every allocated qubit without an operation in a tick has one idle location in that tick, including auxiliary time before preparation and after measurement. Preparation and reset erase an incoming Pauli error. Classical processing is reliable; the verified-cat circuit has one feedback tick, during which data and unused auxiliary qubits idle.

The four-flag circuit prepares five auxiliary qubits, applies H to its measurement qubit, executes its CNOT sequence, and applies H before the Z readout of the measurement qubit. The cat state is grown from five qubits prepared in $|0\rangle$, with H on the root c_0 and the balanced CNOT tree $(c_0, c_1), (c_0, c_2), (c_1, c_3), (c_2, c_4)$. The verified cat checks ports c_1 and c_2 on one verifier before the data interactions and is accepted when the verifier reads 0. The decoded cat uses two checks, on (c_1, c_2) and (c_3, c_4), read after the data interactions; when both are raised it applies X_2X_4 to the row on every block. Cat ports are read in X using five H gates and Z measurements. The four-flag implementation uses five copies of its circuit, with either reused or parallel auxiliary qubits; at $m = 2$ it has 90 CNOTs.

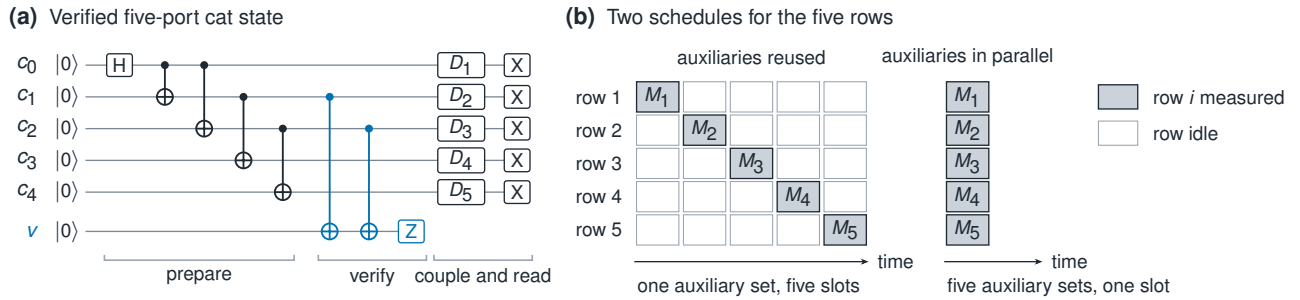


Figure S3: Cat-state preparation and the two schedules. (a) The verified cat checks ports c_1 and c_2 on the verifier v before the data interactions and accepts $v = 0$; D_j couples a port to coordinate j of each block. Each gate has its own column; the second and third CNOTs of the tree share a time step. The decoded cat adds a second verifier on ports c_3 and c_4 and reads both checks after the data interactions. (b) With reused auxiliaries one auxiliary set measures the five rows in succession and each row idles in the other four slots; with parallel auxiliaries five sets measure the rows concurrently. The archived schedules give the gate order and idle positions.

Table S8 gives the location counts: the two-block non-idle totals are 150, 170 and 190, including all Hadamard gates.

Table S8: Two-block, five-row circuits. Prep. counts preparations and resets, Meas. physical Z readouts, and N_{op} their sum with the H gates and CNOTs. Every operation takes one tick. Auxiliaries exclude the 50 data qubits. Verified-cat entries count one accepted attempt.

Circuit	Schedule	Aux.	Prep.	H	CNOT	Meas.	N_{op}	Ticks	Idles	Feedback
Four flags	reused	5	25	10	90	25	150	110	5810	0
Verified cat	reused	6	30	30	80	30	170	65	3390	5
Decoded cat	reused	7	35	30	90	35	190	55	2855	0
Four flags	parallel	25	25	10	90	25	150	22	1410	0
Verified cat	parallel	30	30	30	80	30	170	13	790	1
Decoded cat	parallel	35	35	30	90	35	190	11	655	0

For the verified cat, the failure event is acceptance and logical failure in one attempt, and its coefficients and resource counts refer to one accepted attempt. The reported costs cover the measurement circuits; the output assessment is defined below.

S5.2 Leading-order noise and scheduling

The Shor $[[25, 1, 5]]$ comparison quantifies how idle noise and scheduling determine the relative performance of the four-flag and cat-state circuits. Gates and preparations fail with probability p, using uniform nonidentity Paulis of the appropriate arity. Measurement outcomes flip independently with probability p. Idles fail with probability λp, using uniform nonidentity single-qubit Paulis.

An ideal terminal Shor decoder assesses the output after the measurement circuit. Following the record-dependent corrections, it majority-corrects X errors within each five-qubit row and Z row parities across the five rows of each block. The failure probability P_{fail} counts either a wrong reported joint eigenvalue or a decoded logical Pauli with nontrivial action modulo the single measured product $\prod_b M_b$. The final ideal syndrome and recovery define the assessment; resource counts cover the preceding measurement circuit.

For this assessment the first- and second-order contributions vanish for every circuit, as the enumeration confirms; define the leading coefficient by

$$P_{\mathrm{fail}}(p, \lambda) = c_3(\lambda)p^3 + O(p^4). \tag{S1}$$

Table S9 compares the four-flag, verified-cat and decoded-cat circuits under this common assessment; the time schedules and coefficients are those of the implementations compared here. The ordering changes with idle noise and scheduling. Section S5.4 gives a remainder bound. All 63 evaluations, for 21 configurations at $\lambda = 0, 1/10, 1,$

Table S9: Two-block five-row measurement circuits evaluated with an ideal terminal Shor decoder. Resources count the physical measurement circuit: peak auxiliaries (excluding 50 data qubits), CNOTs and idle locations. Coefficients are in units of 10^3, rounded to three significant digits. Verified-cat resources count one accepted attempt.

Circuit	Schedule	Aux.	CNOTs	Idles	$c_3(0)/10^3$	$c_3(1/10)/10^3$
Four flags	reused	5	90	5810	21.1	2290
Verified cat	reused	6	80	3390	46.4	933
Decoded cat	reused	7	90	2855	72.7	762
Four flags	parallel	25	90	1410	21.1	116
Verified cat	parallel	30	80	790	46.4	149
Decoded cat	parallel	35	90	655	72.7	160

are listed in Table S11, and the 84 polynomial coefficients in Table S10; the archived data contain both as exact rationals.

Each coefficient is an exact cubic polynomial in λ (Fig. S4). With reused auxiliaries the four-flag circuit has the smallest coefficient for $\lambda < \lambda^* \approx 0.022$, and with parallel auxiliaries for $\lambda < \lambda^* \approx 0.158$. Above the corresponding exact crossing, at least one cat-state circuit has a smaller coefficient. The decimal values shown are rounded; exact polynomial differences each have a unique zero in $(0, 1]$.

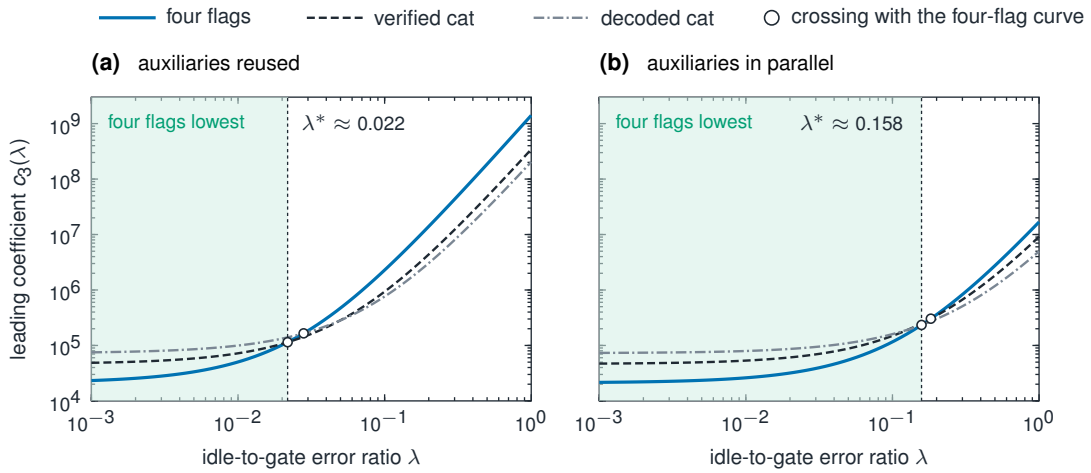


Figure S4: Exact leading coefficient $c_3(\lambda)$ of the two-block five-row readout circuits of Table S9 against the idle-to-gate error ratio λ, with (a) reused and (b) parallel auxiliaries. Each curve is an exact cubic in λ. Open circles mark the crossings of the four-flag curve with the two cat-state curves. The first crossing, at λ^* (dotted line), ends the shaded range in which the four-flag circuit has the smallest coefficient. Verified-cat curves count one accepted attempt.

S5.3 Exact polynomials and crossing intervals

Table S10 lists all 84 rational coefficients of the 21 cubics, and Table S11 their 63 values at $\lambda = 0, 1/10, 1$ together with the location counts that enter $N_{\text{eff}} = N_{\text{op}} + \lambda N_{\text{idle}}$.

The differences between the four-flag and the cat coefficients for two blocks are

$$\Delta_1(\lambda) = -\frac{17065856}{675} - \frac{15986344}{135}\lambda + \frac{979570384}{27}\lambda^2 + \frac{27741845560}{27}\lambda^3,$$
$$\Delta_2(\lambda) = -\frac{6963008}{135} - \frac{44563468}{135}\lambda + \frac{1215926392}{27}\lambda^2 + \frac{10429483960}{9}\lambda^3,$$
$$\Delta_3(\lambda) = -\frac{17065856}{675} - \frac{26739544}{135}\lambda + \frac{13023904}{27}\lambda^2 + \frac{200638720}{27}\lambda^3,$$
$$\Delta_4(\lambda) = -\frac{6963008}{135} - \frac{26008492}{135}\lambda + \frac{44208328}{27}\lambda^2 + \frac{282358960}{27}\lambda^3.$$

Table S10: All 21 exact cubic polynomials, $c_3(\lambda) = \sum_{j=0}^{3} a_j \lambda^j$, with exact coefficients; Table S11 lists their values at three idle ratios. Adaptive denotes the three-then-five stopping rule.

Circuit	Readout	Schedule	m	a_0	a_1	a_2	a_3
Four flags	five votes	reused	1	6727898/675	22244096/45	105473692/9	6073175920/27
Verified cat	five votes	reused	1	18370658/675	2263196/3	25998200/3	243590000/3
Decoded cat	five votes	reused	1	30988498/675	37064576/45	55308112/9	1215838720/27
Four flags	five votes	parallel	1	6727898/675	5243416/45	5267872/9	48437120/27
Verified cat	five votes	parallel	1	18370658/675	775804/3	2569400/3	3110000/3
Decoded cat	five votes	parallel	1	30988498/675	11144176/45	4308512/9	395360
Four flags	adaptive	reused	1	6689498/675	7174168/15	254668972/27	3405682400/27
Four flags	five votes	reused	2	14266066/675	278372492/135	1960971256/27	11996058880/9
Verified cat	five votes	reused	2	10443974/225	98119612/45	327133624/9	8246331080/27
Decoded cat	five votes	reused	2	49081106/675	21529064/9	248348288/9	174063880
Four flags	five votes	parallel	2	14266066/675	12343228/27	98176696/27	343299680/27
Verified cat	five votes	parallel	2	10443974/225	29485228/45	28384264/9	142660960/27
Decoded cat	five votes	parallel	2	49081106/675	29241544/45	17989456/9	60940720/27
Four flags	adaptive	reused	2	69867274/3375	146753668/75	8471395912/135	7860387920/9
Four flags	five votes	reused	3	25533106/675	724582076/135	6857991116/27	45727512400/9
Verified cat	five votes	reused	3	3277022/45	41252224/9	867838244/9	22332373120/27
Decoded cat	five votes	reused	3	72781022/675	24747962/5	675806932/9	4376180440/9
Four flags	five votes	parallel	3	25533106/675	157008596/135	338127128/27	474469120/9
Verified cat	five votes	parallel	3	3277022/45	19518544/15	70586516/9	435011200/27
Decoded cat	five votes	parallel	3	72781022/675	3868246/3	46957780/9	66584920/9
Four flags	adaptive	reused	3	124731274/3375	1157968856/225	1157556952/5	3810185544

Difference	Schedule/comparator	Unique root in $(0, 1]$
Δ_1	reused, verified cat	$(0.021792304333, 0.021792304334)$
Δ_2	reused, decoded cat	$(0.028015257547, 0.028015257548)$
Δ_3	parallel, verified cat	$(0.182771449952, 0.182771449953)$
Δ_4	parallel, decoded cat	$(0.157904387208, 0.157904387209)$

Each interval contains the unique root of its polynomial in $(0, 1]$, certified with Sturm sequences in exact rational arithmetic.

Near a crossing, higher-order contributions determine the finite-p ordering. Section S5.4 bounds the absolute asymptotic error for fixed circuits.

Freund et al. use pure-Z dephasing with $q_Z = [1 - \exp(-\Delta t/T_2)]/2$. Their $\Delta t = 350\,\mu$s, $T_2 = 200\,$ms and two-qubit error probability 0.02 give the rough ratio $q_Z/0.02 \simeq 0.044$. The present model uses isotropic Pauli idle noise and equal-duration operations; the experiment also has unequal gate, preparation and measurement durations and measurement-induced noise. This estimate motivates the idle-noise range studied here.

S5.4 A finite-circuit remainder bound

For the failure event defined in Sec. S5.2, consider a fixed finite measurement circuit with independent location rates $r_i p$, and put $N_{\text{eff}} = \sum_i r_i = N_{\text{op}} + \lambda N_{\text{idle}}$, where N_{op} counts non-idle physical locations (preparations, gates, and measurements) and N_{idle} counts idle qubit-ticks. When no event of order at most two fails, a direct union bound gives

$$\begin{aligned} \max\{0, (1 - N_{\text{eff}} p) c_3 p^3\} &\le P_{\text{fail}} \\ &\le \min\{1, c_3 p^3 + N_{\text{eff}}^4 p^4/24\}, \qquad \text{(S2)} \\ |P_{\text{fail}} - c_3 p^3| &\le (c_3 N_{\text{eff}} + N_{\text{eff}}^4/24) p^4. \qquad \text{(S3)} \end{aligned}$$

Equation (S2) bounds the absolute error for the fixed four-flag and decoded-cat schedules. Its useful perturbative range depends on the displayed coefficients and N_{eff}. The adaptive three-then-five controller measures three representatives and, on disagreement, adds the remaining two before taking the majority. The archive also supplies the leading coefficients of this adaptive controller.

Table S11: Exact leading coefficients $c_3(\lambda)$ of the 21 configurations at $\lambda = 0$, 1/10 and 1, with the numbers N_{op} and N_{idle} of non-idle and idle locations; $N_{\mathrm{eff}} = N_{\mathrm{op}} + \lambda N_{\mathrm{idle}}$. The verified cat counts one accepted attempt and the adaptive readout its longer branch.

Circuit	Readout	Schedule	m	N_{op}	N_{idle}	$c_3(0)$	$c_3(1/10)$	$c_3(1)$
Four flags	five votes	reused	1	125	2360	6727898/675	30114301/75	17786701582/75
Verified cat	five votes	reused	1	145	1660	18370658/675	182596268/675	61184934758/675
Decoded cat	five votes	reused	1	165	1370	30988498/675	158462414/675	35131033538/675
Four flags	five votes	parallel	1	125	660	6727898/675	19754854/675	1691397538/675
Verified cat	five votes	parallel	1	145	460	18370658/675	42307148/675	1470791558/675
Decoded cat	five votes	parallel	1	165	370	30988498/675	51203014/675	788157538/675
Four flags	adaptive	reused	1	125	2390	6689498/675	187782557/675	91838311358/675
Four flags	five votes	reused	2	150	5810	14266066/675	57162946/25	35190178738/25
Verified cat	five votes	reused	2	170	3390	10443974/225	126003967/135	232196424902/675
Decoded cat	five votes	reused	2	190	2855	49081106/675	514303421/675	137783001506/675
Four flags	five votes	parallel	2	150	1410	14266066/675	78250802/675	11359756166/675
Verified cat	five votes	parallel	2	170	790	10443974/225	100414486/675	6168954142/675
Decoded cat	five votes	parallel	2	190	655	49081106/675	11995448/75	373381274/75
Four flags	adaptive	reused	2	150	5865	69867274/3375	5795753228/3375	3166104150134/3375
Four flags	five votes	reused	3	175	10510	25533106/675	5531885353/675	3604661651386/675
Verified cat	five votes	reused	3	195	5370	3277022/45	1567735021/675	125308053686/135
Decoded cat	five votes	reused	3	215	4590	72781022/675	1241947241/675	382312808792/675
Four flags	five votes	parallel	3	175	2410	25533106/675	44830874/135	44848938286/675
Verified cat	five votes	parallel	3	195	1170	3277022/45	40160789/135	3419351702/135
Decoded cat	five votes	parallel	3	215	990	72781022/675	200028761/675	9458838872/675
Four flags	adaptive	reused	3	175	10590	124731274/3375	4506914039/675	13658221417714/3375

S6 Enumeration and independent verification

S6.1 Finite implementation

Independent forward-Pauli and reverse-image implementations intersect the admissible corrections for each record, modulo the measured operator. A separate scalar checker uses every single-fault image and a fixed correction table, without reducing the fault list or solving correction intersections. The quotient weight is the residual weight minimized over multiplication by the measured operator, using X weight or full Pauli weight for the certificate being tested. Removing an unflagged image of quotient weight at most one is justified by the lower-order constraint and the triangle inequality; repeated-location factors combine by Pauli closure. Complete fixed-table checks combine single-fault images into effect multisets, unordered collections allowing repetitions. At distance eleven they cover $1,183,009,464$ order-five effect multisets per boundary order, with maximum residual weights $0, 1, 2, 3, 4, 5$. These counts enumerate Pauli-effect multisets.

S6.2 Coverage and independent comparisons

The exhaustive collapse comparisons use the parameter triples below. Additional random samples cover $t = 1, w = 3 \ldots 8$; $t = 2, w = 5 \ldots 8$; $t = 3, w = 7 \ldots 9$.

t	1	1	1	1	2	2	2
w	3	4	5	6	5	6	7
k	1	2	3	2	2	3	2

Additional cat and hybrid cases occur both on and above the boundary. For every tested record, the admissible correction sets for X_t and L_t agree above the boundary; those for X_t and L_{t-1} agree at $w = 2t + 1$. Samples check the finite implementations.

The zero-flag guarantee of Corollary 6 is checked on the sixteen two-block orders of the two-flag circuit $D_1F_1D_2F_2D_3D_4F_1D_5F_2$ of Section S1, accepted at $f = 0$. Heterogeneous lifting is checked on six support choices, four deterministic random orders per choice, including empty supports and three blocks. Ideal Kraus semantics is checked by integer matrix columns on twenty-four GHZ-tree examples with intermediate flag couplings. The two

even-coupling variants of Anker–Marvian Fig. 8 (Section S7.1) are separately checked on all sixteen orders. Small counterexample records and the two-fault physical history are retained.

S6.3 Resource classification

Table S12 summarizes the flag-count bounds for circuits with one measurement qubit and twice-coupled flags.

Table S12: Minimum flag counts for circuits with one measurement qubit and twice-coupled flags. Deterministic corrections accept every reachable record; a zero-flag guarantee constrains the zero record. Values outside brackets are established by exhaustive exclusions and checked examples. Brackets denote an interval for an unresolved minimum.

$d = w$	t	Certificate	X-weight	Pauli-weight
3	1	all records	0	1
3	1	zero flag	0	1
5	2	all records	2	4
5	2	zero flag	1	2
7	3	all records	6	$[6, 9]$
7	3	zero flag	3	4

Allowing two, four, or six couplings per flag leaves the $d = 7$ correction minima in the X-weight interval $[5, 6]$ and the Pauli-weight interval $[5, 9]$. The five-flag mixed-coupling cell is unresolved.

The wider class $\mathcal{C}_2$ has c GHZ ports and k flags, arbitrary birth-ordered GHZ trees, surjective port assignments, active-port flag couplings during preparation and between data layers, terminal measurements, and exactly two couplings per flag. Flags are targets only and repeated identical endpoints are excluded. The class $\mathcal{C}_{2,4,6}$ instead permits two, four, or six couplings per flag. These search classes are smaller than the class of Theorem 2. At $w = 5, t = 2$, five auxiliary qubits are necessary in $\mathcal{C}_2$, and the only feasible type with five is $(c, k) = (1, 4)$. In $\mathcal{C}_{2,4,6}$, the CNOT minimum is $5m + 8$, and the minimum number of auxiliary qubits is four or five. The mixed-coupling $(2, 2)$ cell remains unresolved.

Birth-ordered GHZ trees enumerate every new port's parent among earlier ports. Active-port flag endpoints include preparation slots and all data boundaries; the assignment of data coordinates to ports is surjective. Within-boundary permutations are included. The enumeration counts include redundant endpoint representations; the supplied programs regenerate the algebraic exclusion certificates. Table S13 lists the principal resource classes and their coverage.

Table S13: Selected resource-classification rows at $w = 5, t = 2$. All counts except the final unresolved row come from exhaustive enumeration. Counts use endpoint representations, which can represent the same circuit more than once.

c	k	Class	Candidates	Feasible representations
1	3	$\mathcal{C}_2$	3,955	0
2	2	$\mathcal{C}_2$	179,640	0
3	1	$\mathcal{C}_2$	68,700	0
4	0	$\mathcal{C}_2$	1,440	0
1	4	$\mathcal{C}_2$	84,540	3,453
2	3	$\mathcal{C}_2$	13,172,640	0
3	2	$\mathcal{C}_2$	12,550,500	0
4	1	$\mathcal{C}_2$	684,000	0
5	0	$\mathcal{C}_2$	2,880	0
2	2	$\mathcal{C}_{2,4,6}$	8,710,029,240	unresolved

For the $\mathcal{C}_{2,4,6}$ CNOT bound, write $N_{\mathrm{CNOT}} = 5m + D$ with $D = c - 1 + H$ and $H \geq 2k$, where H is the total number of port-to-flag CNOTs. If $D < 8$, then $c + 2k \leq 8$ and, since $c \leq 5$, at most six auxiliary qubits need consideration. The exhaustive budget exclusions cover all types with at most five. With six, only $(c, k) = (4, 2), (5, 1)$ can improve the gate bound, and both are fully excluded in the two-coupling subclass, which already exhausts their available gate budget. The four-flag circuit attains $5m + 8$.

S6.4 Classification at fixed layer orders and physical cores

The first sequence enumerator inserts flag endpoints recursively while enforcing their first-appearance labels. The second selects a multiset of k pairs from the fifteen unordered boundary pairs in $\{0,\dots,5\}$, permutes the incident flags independently at every boundary, and normalizes labels by first appearance. Every sequence of the family has such an endpoint multiset and slot permutations; normalization and a set remove exactly the redundant descriptions. The two enumerators give identical sorted sequence lists for $k = 0, 1, 2, 3$, giving $1, 15, 195, 2605$ sequences. Their sixteen layer orders give 45,056 cases.

For a CNOT sequence with one measurement qubit, an X on that qubit at a cut produces the suffix of its data and flag targets. A target X produces a unit vector; a correlated gate X fault also permits their XOR. These are precisely the X projections of the full single-location Pauli alphabet: Z factors leave every X component and every measured Z flag unchanged. All single-qubit cuts are included. Zero, one, and two such effects therefore cover all events through two faults. The separate implementation propagates X columns backwards. Same-location factors can be multiplied and identities removed. If this lowers a two-factor event's cost, its constraint at bound two is weaker than the physical lower-cost constraint, so every exclusion remains valid.

Each event imposes one union of two balls on (r_A, r_B), one for each common choice. The independent checker retains all 1024 pairs; the second uses 512 classes under simultaneous complementation. A separate forward interpreter verifies the 298 physical events in 96 inclusion-minimal conflict cores for the two small circuits. A conflict core is a set of same-record events with no common admissible correction pair; inclusion-minimality means that removing any event restores a nonempty intersection. The four-flag positive control passes all sixteen X-weight tests; its Pauli-weight certificate is supplied by the independent fixed-table check, and ideal G0 follows analytically from even flags.

S6.5 Readable obstruction certificates

The 96 cores give short obstruction witnesses for two circuits across all sixteen two-block orders; the enumeration in Section S6.4 separately covers 45,056 cases.

A three-event core can be checked by hand. Use

$$F_2 D_1 D_2 F_1 D_3 D_4 F_2 D_5 F_1$$

with block A before B in every data layer, and write a for the measurement qubit. All three events below produce the record $(f_1, f_2) = (0, 1)$:

Cost	Fault insertions	A residual	B residual
1	X on flag 2 after preparation	I	I
1	X_a after the data gate to B_2	$X_3X_4X_5$	$X_3X_4X_5$
2	X_a after the gate to A_3 and after the closing F_2	X_4	X_3X_4

Suppose any two X corrections r_A, r_B satisfy the graded bound with one common choice. By the first event, simultaneous complementation lets us take each correction to have weight at most one. The second event must then use $g = 1$, and forces each correction to be either X_1 or X_2. For the third event, $g = 0$ leaves weight three on B and $g = 1$ leaves weight three on A. This is a contradiction. Removing any one event makes the constraints feasible, as verified by enumerating all $32^2 = 1024$ correction pairs.

A higher-order obstruction. Prabhu's Fig. 2.8(b) [9] gives the weight-seven sequence

$$F_1F_2D_1F_3D_2F_4D_3F_5D_4F_1D_5F_6D_6F_3D_7F_5F_2F_4F_6.$$

It has a Pauli-weight table through two faults, consistent with its distance-five design. For this comparison we choose our own table: the smallest integer admissible correction at each record, with coordinate 1 the least significant bit. This gives $r = 0$ at record $(1, 0, 1, 0, 0, 0)$; records first reached at three faults may use any correction. The same table satisfies the three-fault X-weight test because every seven-bit X string has a representative of weight at most three.

Insert X on flag 6 after preparation, X_aY_4 after D_4, and X_aZ_7 after D_7. The record is $(1, 0, 1, 0, 0, 0)$ and the residual is $Y_4X_5X_6Y_7$. Its representatives have weights four and five, so this table fails the three-fault Pauli bound. A four-event core in the archive also proves that no X correction works at this record through three faults. For a shared witness, order layer 4 as B then A and all others as A then B. Keep the initial flag fault, but use X_a

after the data gates to B_4 and A_7. The same record gives $X_4X_5X_6X_7$ on A and X_5X_6 on B, with weights $(4, 2)$ and $(3, 5)$ for the two common choices.

S6.6 Independent propagation checks

The general statements are proved in Appendix A of the main text. Direct finite checks also compare independently implemented propagation maps. A binary-matrix implementation tests all seven weight-three sequences with zero or one flag on two and three blocks, covering 280 lifts modulo a global block permutation. For the weight-five four-flag circuit it tests all sixteen representative two-block orders and six sampled three-block orders through two faults. The six three-block orders are a sample of that circuit's order space. For the mixed-Pauli Freund circuit, all eight two-block and all 216 three-block orders pass the one-fault check.

Stim [6] supplies a separate stabilizer implementation. Independently transcribed primitive gates are checked on one, two and three blocks, with every single-location Pauli and 768 distinct-location two-fault samples for the weight-five circuit. Encoded-reference tests check the ideal product instrument, and a separate full-protocol test covers the Steane branches in Section S4. The forward replay checks the 298 physical events and all 1024 correction pairs in the 96 conflict cores.

S7 Comparison with earlier constructions

Table S14 compares measurement guarantees and recovery information. The record lists the classical outcomes available to each recovery procedure. Here t is the certified fault order, $w \geq 2$ is the single-block support weight, f is the Z-basis auxiliary record, and M_b is the measured Pauli on block b. For a fixed record and X correction, X_t bounds the support X weight by the fault count through order t; L_t bounds the full residual Pauli weight. Both minimize over the choice g. After sharing, one common choice $g \in \{0, 1\}$ multiplies every block by M_b^g, while the applied correction depends on f alone. Define $q = \min\{t, \lfloor w/2 \rfloor - 1\}$. The table uses Calderbank–Shor–Steane (CSS) codes, quantum error correction (QEC) and linear programming (LP).

Table S14: Measurement guarantees and recovery information in earlier constructions and this work. Section and theorem numbers refer to the versions listed in the references.

Work	Measured operator	Record and recovery rule	Guarantee	Construction or transfer statement
Chamberland–Beverland [10]	Weight-w Pauli	Flags; error-correction protocols additionally use syndromes	Definition 6: excessive full Pauli weight modulo measured Pauli implies a flag	Every set of up to t faults; ideal projective and zero-flag conditions included
Chao–Reichardt [11]	Arbitrary stabilizer	Flags determine correction; stabilizer readout retained	Full residual-weight control at arbitrary distance	Definition 2, padded Lemma 4; transferred here by main-text Corollary 11
Prabhu–Reichardt [12]	Stabilizer and cat preparation	Flag record in listed correction rules	Separate CSS X/Z weight criterion; higher-distance gadgets	Explicit distance-five/seven gadgets; listed code-dependent guarantees
Tansuwannont et al. [13]	Logical Pauli, including multiple blocks	Flags, repeated outcomes, later data syndromes	Fault-tolerant measurement for specified cyclic CSS codes	Distance three; cross-block argument in Section V, computation in VI
Anker–Marvian (2024) [2]	Stabilizers and connected syndrome sequences	Flags plus a syndrome history to adjust	Full-Pauli strong-fault-tolerance definition; binary matrix correction	Classical-code construction; connected syndromes (App. C), paired data gates (App. D)
Anker–Marvian (2025) [14]	General Clifford circuits	Circuit-dependent flags, meta-flags; front-end error correction	Fault-tolerant construction for prescribed Clifford circuits	Constructive Clifford protection with redesigned flag circuitry

Table S14 (continued)

Work	Measured operator	Record and recovery rule	Guarantee	Construction or transfer statement
Liou–Lai [15]	Multiple stabilizer measurements	Shared flags and generalized syndrome, including follow-up checks	Fault-tolerant extraction under their distinguishability conditions	Lemmas 6/10 and Proposition 9 preserve local CNOT order and constrain allowed simultaneous extraction
Ghosh et al. [16]	General circuit; Appendix C stabilizer gadget	Retained record; ideal recovery also accesses quantum output	Exact noiseless recovery (Thm. IV.3); gadget distance upper bounds (App. C)	Circuit-specific noise set; flag-LP infeasibility bounds distance
Benhemou–Berthusen [17]	Inter-code Clifford gates and Pauli products	Code checks and additional flag outcomes	Flags restore circuit distance in specified gadgets	Chain-map constructions and examples; complete measurement can need intervening QEC or a logical flag
This work	Shared product $\prod_b M_b$	Z-basis record only for correction/control	Blockwise residual weight with one common choice	Local Pauli-weight certificate iff all two-block X-weight tests; fixed-lift projection; $\mathsf{X}_t \iff \mathsf{L}_q$

Anker–Marvian's connected-syndrome construction tracks corrections through the syndrome history and delays flag readout; its resource comparison assumes slow or absent reset. Appendix C follows the residual after flag correction. Appendix D places two data CNOTs in each flagged region and applies the correction between them, using fault pairing across at least $2t+1$ parity-check repetitions. Their general-Clifford construction [14] designs flags and meta-flags together with front-end error correction.

Ghosh et al. Definition IV.1 and Theorem IV.3 (arXiv v3) formulate exact channel recovery by a noiseless map receiving the retained record and quantum output. Definition C.8, Eq. (C17), gives a Pauli-weight zero-flag condition modulo the measured Pauli; Theorem C.13 gives an LP distance upper bound for their circuit family. Benhemou–Berthusen, Section 7.1, simulate flags using ideal Pauli-product measurements with noisy reported bits.

The present criterion characterizes reuse of a given auxiliary circuit and correction table: local full-Pauli protection is equivalent to protection across every block count and layer order, with necessity witnessed on two blocks.

S7.1 Source-circuit conventions

The two-flag weight-five sequence in Section S1 is one of two even-coupling variants of Anker–Marvian Fig. 8 [2] that realize the corrections in their Table VI. The printed arXiv v2 and journal figure has the sequence

$$F_2\ D_1\ F_2\ D_2\ F_1\ D_3\ D_4\ F_2\ D_5\ F_1,$$

with three F_2 couplings. Under that paper's general preparation convention, the measurement qubit starts in $|+\rangle$ and the flags in $|0\rangle$; terminal measurements are X and Z. Let $v = f_2 \in \{0,1\}$ be the Z readout of the flag with three couplings, and let $y \in \{0,1\}$ be the X readout of the measurement qubit. For $f_1 = 0$, direct branch evaluation gives $K_{v,y} = (-1)^{vy} M^v/2$; the branches with $f_1 = 1$ vanish. Deleting either of the first two F_2 gates gives an even-coupling sequence with G0. Both variants admit the four corrections in their Table VI and the corrections admitted by X_2 listed in Section S1, and neither admits an L_2 correction. They provide concrete single-block X_2 certificates whose sharing is tested here.

Reading the literature definitions

Anker–Marvian's Definition 3 includes a full-Pauli strong-fault-tolerance condition. Its Eq. (2) [2] constrains the binary X-error and flag syndrome. For a serial circuit with one measurement qubit, correlated X factors at a flag or data CNOT can be moved to the opposite cut on that qubit at equal or lower cost. Pure direct-data factors are restored by a triangle inequality. This connects the formula to our X-weight certificate when its column set includes the required cuts. We independently check the two even-coupling variants in the full physical model.

Prabhu–Reichardt define CSS fault tolerance using separate X and Z weights [12], Section 2. Their Appendix C, Figs. 13–15, includes distance-five and distance-seven circuits, also discussed in Prabhu's dissertation, Section 2.3 [9]. The three dissertation Fig. 2.8 sequences are checked here with all Pauli gate faults and cut faults, including all idle-fault images. Their weights $6, 7, 8$ exceed the $t = 2$ boundary. Chamberland–Beverland Definition 6 already uses full Pauli weight and zero-flag guarantees, as discussed in Section III. The different residual-weight conditions are compared in Table S15.

Table S15: Gate sequences under the full Pauli fault model. Pass and fail refer to exhaustive fault checks. A passing Pauli-weight certificate supplies the shared blockwise bound. The verified cat is restricted to its accepted branch.

Circuit	w	t	Boundary?	X-weight	Pauli-weight	Sharing
Prabhu, Fig. 2.8(a–c)	6,7,8	2	no	pass	pass	yes
Two-flag weight-five example	5	2	yes	pass	fail	no
Five-flag weight-six sequence	6	2	no	pass	pass	yes
Verified cat, accepted	5	2	yes	pass	pass	accepted
Decoded cat	5	2	yes	pass	pass	yes

Table S14 compares the recovery procedures of these and further constructions.

The weight-seven circuit in Prabhu's Fig. 2.8(b) meets its original two-fault target. Section S6.5 gives the physical obstruction at the higher three-fault target.

S7.2 Resources for joint measurement

Table S16 compares resources and recovery information for a single joint Pauli measurement.

Table S16: Resources and recovery for one joint Pauli measurement. $W = mw$ is its total support and h is the selected Chao–Reichardt padding.

Construction	Primitive cost or requirement	Recovery information and output
Shared Chao–Reichardt	$W + 4t(w + h)$ CNOTs; $2t + 2$ auxiliaries	Flag-only correction; at most s residual per block with one common product representative
Direct Chao–Reichardt	$W + 4t(W + h)$ CNOTs; $2t + 2$ auxiliaries	Flag-only correction; at most s residual on the whole measured support
Verified/decoded cat	Cat ports and verification/decoding; preparation may retry	Verification outcomes or decoded-cat record; protocol must protect the logical report and output
Transversal/encoded ancilla	Suitable transversal coupling and an encoded ancilla	Requires code-compatible logical coupling and fault-tolerant ancilla preparation
Lattice surgery	Compatible local checks and repeated boundary syndrome rounds	Uses geometry, check history and a joint decoder; cost depends on distance and layout

The cat methods originate in Refs. [7, 8]; surgery methods are discussed in Refs. [18, 19]. The Steane comparison in Section S4 fixes the code, syndrome extraction, timing and record, so that the bare circuit receives the same opportunity to use its full measurement history.

References

[1] A. Chen, Code and data for "Sharing flag circuits across code blocks: an exact local criterion for joint logical measurements", Zenodo (2026), https://doi.org/10.5281/zenodo.24203932.

[2] B. Anker and M. Marvian, PRX Quantum **5**, 040340 (2024), arXiv:2212.10738v2.

[3] R. Freund, F. Butt, C. Benito, I. Pogorelov, M. Meyer, A. Steiner, A. Bermudez, M. Müller, and T. Monz, Modular fault-tolerant quantum computing on a non-CSS code (2026), arXiv:2609.22572v1 [quant-ph].

[4] R. Freund *et al.*, Modular fault-tolerant quantum computing on a non-CSS code, Zenodo (2026), https://doi.org/10.5281/zenodo.22796458.

[5] R. Chao and B. W. Reichardt, Phys. Rev. Lett. **121**, 050502 (2018), arXiv:1705.02329v1.

[6] C. Gidney, Quantum **5**, 497 (2021).

[7] P. W. Shor, in *Proceedings of the 37th Annual Symposium on Foundations of Computer Science* (IEEE, 1996) pp. 56–65, arXiv:quant-ph/9605011v2.

[8] D. P. DiVincenzo and P. Aliferis, Phys. Rev. Lett. **98**, 020501 (2007), arXiv:quant-ph/0607047v2.

[9] P. Prabhu, *Lower overhead fault-tolerant building blocks for noisy quantum computers*, Ph.D. thesis, University of Southern California (2024), arXiv:2605.12385v1.

[10] C. Chamberland and M. E. Beverland, Quantum **2**, 53 (2018), arXiv:1708.02246v3.

[11] R. Chao and B. W. Reichardt, PRX Quantum **1**, 010302 (2020), arXiv:1912.09549v1 [quant-ph].

[12] P. Prabhu and B. W. Reichardt, Quantum **7**, 1154 (2023), arXiv:2108.02184v2.

[13] T. Tansuwannont, C. Chamberland, and D. Leung, Phys. Rev. A **101**, 012342 (2020), arXiv:1803.09758v3.

[14] B. Anker and M. Marvian, Universal fault tolerance with non-transversal Clifford gates (2025), arXiv:2510.08402v1.

[15] P.-H. Liou and C.-Y. Lai, IEEE Trans. Quantum Eng. **6**, 3101524 (2025), arXiv:2407.00607v1.

[16] S. Ghosh, H. Boche, and A. Tanggara, Fault tolerance of quantum circuits with tensor networks and symplectic geometry (2026), arXiv:2610.03429v3.

[17] A. Benhemou and N. Berthusen, Automated logical Clifford gadgets for heterogeneous architectures via chain maps (2026), arXiv:2607.02482v1.

[18] D. Horsman, A. G. Fowler, S. Devitt, and R. Van Meter, New J. Phys. **14**, 123011 (2012), arXiv:1111.4022.

[19] D. Litinski, Quantum **3**, 128 (2019), arXiv:1808.02892.