Density-matrix simulation of small surface codes under current and projected experimental noise

T. E. O'Brien[*],[1] B. Tarasinski[*],[1, 2] and L. DiCarlo[2, 3]

[1]*Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands*[*]
[2]*QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands*
[3]*Kavli Institute of Nanoscience, Delft University of Technology,*
P.O. Box 5046, 2600 GA Delft, The Netherlands
(Dated: April 7, 2017)

We present a full density-matrix simulation of the quantum memory and computing performance of the distance-3 logical qubit Surface-17, following a recently proposed quantum circuit and using experimental error parameters for transmon qubits in a planar circuit QED architecture. We use this simulation to optimize components of the QEC scheme (e.g., trading off stabilizer measurement infidelity for reduced cycle time) and to investigate the benefits of feedback harnessing the fundamental asymmetry of relaxation-dominated error in the constituent transmons. A lower-order approximate calculation extends these predictions to the distance-5 Surface-49. These results clearly indicate error rates below the fault-tolerance threshold of surface code, and the potential for Surface-17 to perform beyond the break-even point of quantum memory. At state-of-the-art qubit relaxation times and readout speeds, Surface-49 could surpass the break-even point of computation.

I. INTRODUCTION

Recent experimental demonstrations of small quantum simulations [1–3] and quantum error correction (QEC) [4–7] position superconducting circuits for targeting quantum supremacy [8] and quantum fault tolerance [9], two outstanding challenges for all quantum information processing platforms. On the theoretical side, much modeling of QEC codes has been made to determine fault-tolerance threshold rates in various models [10–12] with different error decoders [13–15]. However, the need for computational efficiency has constrained many previous studies to oversimplified noise models, such as depolarizing and bit-flip noise channels. This discrepancy between theoretical descriptions and experimental reality compromises the ability to predict the performance of near-term QEC implementations, and offers limited guidance to the experimentalist through the maze of parameter choices and trade-offs. In the planar circuit quantum electrodynamics (cQED) [16] architecture, the major contributions to error are transmon qubit relaxation, dephasing from flux noise and resonator photons leftover from measurement, and leakage from the computational space, none of which are well-approximated by depolarizing or bit-flip channels. Simulations with more complex error models are now essential to accurately pinpoint the leading contributions to the logical error rate in the small-distance surface codes [10, 13, 17] currently pursued by several groups worldwide.

In this paper, we perform a full density-matrix simulation of the distance-3 surface code named Surface-17, using the concrete quantum circuit recently proposed in [18] and the measured performance of current experimental

multi-transmon cQED platforms [19–22]. For this purpose, we have developed an open-source density-matrix simulation package named quantumsim [23]. We use quantumsim to extract the logical error rate per QEC cycle, $\epsilon_{\rm L}$. This metric allows us to optimize and trade off between QEC cycle parameters, assess the merits of feedback control, predict gains from future improvements in physical qubit performance, and quantify decoder performance. We compare an algorithmic decoder using minimum-weight perfect matching (MWPM) with home-made weight calculation to a simple look-up table (LT) decoder, and weigh both against an upper bound (UB) for decoder performance obtainable from the density-matrix simulation. Finally, we make a low-order approximation to extend our predictions to the distance-5 Surface-49. The combination of results for Surface-17 and -49 allows us to make statements about code scaling and to predict the code size and physical qubit performance required to achieve break-even points for memory and computational performance.

II. RESULTS

A. Error rates for Surface-17 under current experimental conditions

To quantify the performance of the logical qubit, we first define a test experiment to simulate. Inspired by the recent experimental demonstration of distance-3 and -5 repetition codes [4], we first focus on the performance of the logical qubit as a quantum memory. Specifically, we quantify the ability to hold a logical $|0\rangle$ state, by initializing this state, holding it for $k \in \{1, \ldots, 20\}$ cycles, performing error correction, and determining a final logical state (see Fig. 6 for details). The logical fidelity $\mathcal{F}_{\rm L}[k]$ is then given by the probability to match the ini-

tial state. We observe identical results when using $|1\rangle$ or $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ in place of $|0\rangle$.

Parameter	Symbol	Value	Reference
Qubit relaxation time	T_1	30 μs	[19]
Qubit dephasing time (white noise)	T_ϕ	60 μs	[19]
Single-qubit gate time	$\tau_{\mathrm{g,1Q}}$	20 ns	[19, 21]
Two-qubit gate time	$\tau_{\mathrm{g,2Q}}$	40 ns	[5]
Coherent step time	τ_{c}	200 ns	[18]
Measurement time	τ_{m}	300 ns	[19]
Depletion time	τ_{d}	300 ns	[19]
Fast measurement time	$\tau_{\mathrm{m}}^{\mathrm{(fast)}}$	100 ns	[22]
Fast depletion time	$\tau_{\mathrm{d}}^{\mathrm{(fast)}}$	100 ns	[22]

TABLE I. Summary of standard times used in all density-matrix simulations, unless otherwise indicated. The two-qubit gate is a conditional phase gate (C-Z). Other error rates and parameters are given in App. B.

We base our error model for the physical qubits on current typical experimental performance for transmons in planar cQED, using parameters from the literature and in-house results (e.g., gate-set tomography measurements). These are summarized in Table I, and further detailed in App. B. We focus on the QEC cycle proposed in [18], which pipelines the execution of X- and Z-type stabilizer measurements. Each stabilizer measurement consists of three parts: a coherent step (duration $\tau_{\mathrm{c}} = 2\tau_{\mathrm{g,1Q}} + 4\tau_{\mathrm{g,2Q}}$), measurement ($\tau_{\mathrm{m}}$), and photon depletion from readout resonators (τ_{d}), making the QEC cycle time $\tau_{\mathrm{cycle}} = \tau_{\mathrm{c}} + \tau_{\mathrm{m}} + \tau_{\mathrm{d}}$.

Simulating this concrete quantum circuit with the listed parameters using quantumsim, we predict $\mathcal{F}_{\mathrm{L}}[k]$ of Surface-17 (Fig. 1). We show $\mathcal{F}_{\mathrm{L}}[k]$ for both a home-made MWPM decoder (green, described in App. C), and an implementation of the LT decoder of [13] (blue, described in App. D). To isolate decoder performance, we can compare the achieved fidelity to an upper bound extractable from the density-matrix simulation (red, described in Sec. IV A 3). To assess the benefit of QEC, we also compare to a single decohering transmon, whose fidelity is calculated by averaging over the six cardinal points of the Bloch sphere:

$$\mathcal{F}_{\mathrm{phys}}(t) = \frac{1}{6}\left(1 + e^{-t/T_1}\right) + \frac{1}{3}\left(1 + e^{-t(1/2T_1 + 1/T_\phi)}\right).$$
(1)

The observation of $\mathcal{F}_{\mathrm{L}}[k] > \mathcal{F}_{\mathrm{phys}}(k\tau_{\mathrm{cycle}})$ for large k would constitute a demonstration of QEC beyond the quantum memory break-even point [7]. Equivalently, one can extract a logical error rate ϵ_{L} from a best fit to $\mathcal{F}_{\mathrm{L}}[k]$ (see Sec. IV A 2), and compare to the physical error rate

$$\epsilon_{\mathrm{phys}} = -\tau_{\mathrm{cycle}} \left.\frac{d\mathcal{F}_{\mathrm{phys}}(t)}{dt}\right|_{t=0} = \frac{\tau_{\mathrm{cycle}}}{3T_1} + \frac{\tau_{\mathrm{cycle}}}{3T_\phi}.$$
(2)

We observe $\epsilon_{\mathrm{L}} = 1.44\%_{\mathrm{c}}$ for the LT decoder, $\epsilon_{\mathrm{L}} = 1.07\%_{\mathrm{c}}$ for the MWPM decoder, and $\epsilon_{\mathrm{L}} = 0.68\%_{\mathrm{c}}$ at the decoder upper bound ($\%_{\mathrm{c}} = \%$ per cycle). The

latter two fall below $\epsilon_{\mathrm{phys}} = 1.33\%_{\mathrm{c}}$. Defining the decoder efficiency $\eta_{\mathrm{d}} = \epsilon_{\mathrm{L}}^{\mathrm{(UB)}}/\epsilon_{\mathrm{L}}$, we find $\eta_{\mathrm{d}}^{\mathrm{(LT)}} = 0.47$ and $\eta_{\mathrm{d}}^{\mathrm{(MWPM)}} = 0.64$.

We can also compare the multi-cycle error correction to majority voting, in which the state declaration is based solely on the output of the final data qubit measurements (ancilla measurements are ignored). Majority voting corrects any single data qubit error (over the entire experiment), and thus exhibits a quadratic decay for small k [24]. A decoder should also be able to correct (at least) a single error, and thus should produce the same behavior at low k, delaying the onset of exponential decay in $\mathcal{F}_{\mathrm{L}}[k]$. In fact, a good test for the performance of a MWPM decoder is to ensure it can outperform the majority vote at short timescales, as suboptimal configuration will prevent this (as seen for the look-up table decoder).
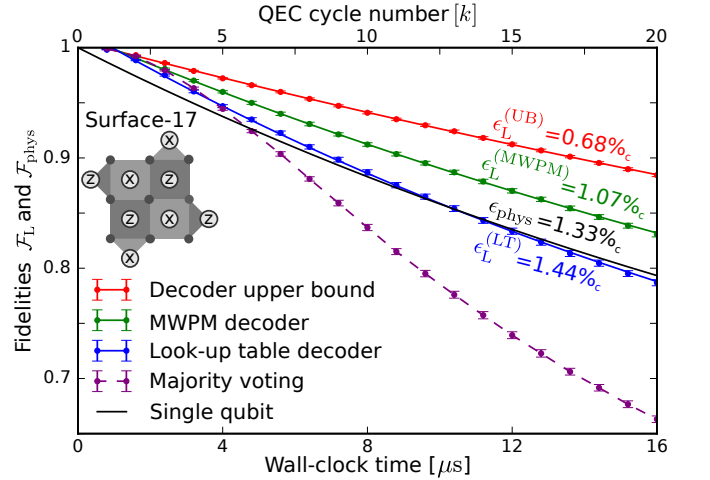

FIG. 1. Logical fidelity $\mathcal{F}_{\mathrm{L}}[k]$ of Surface-17 with current experimental parameters (Table I and App. B), simulated with quantumsim as described in Fig. 6. The results from a MWPM decoder (green) and an implementation of the LT decoder of [13] (blue) are compared to the decoder upper bound (red). The labeled error rate is obtained from the best fit of a compensated exponential decay (also plotted). A further comparison is given to majority voting (purple, dashed), which ignores the outcome of individual stabilizer measurements, and to the fidelity $\mathcal{F}_{\mathrm{phys}}$ of a single transmon (black) [Eq. (1)]. Error bars are obtained by bootstrapping.

With the baseline for current performance established, we next investigate ϵ_{L} improvements that may be achieved by two means. First, we consider modifications to the QEC cycle at fixed physical performance. Afterwards, we consider the effect of improving physical qubit T_1 and T_ϕ.

B. Optimization of logical error rates with current experimental conditions

Error sources in current cQED setups derive primarily from transmon decoherence, as opposed to gate and mea-
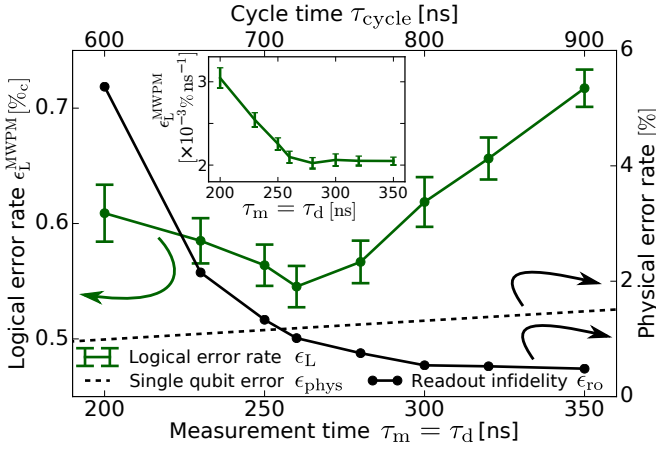
FIG. 2. Optimization of the logical error rate (per cycle) of Surface-17 as a function of measurement-and-depletion time [19]. Changes in the underlying physical error rates are shown as well. Decreasing the measurement time causes an increase in the readout infidelity (solid black curve with dots), whilst decreasing the single qubit decay from T_1 and T_2 (black dashed curve) for all qubits. The logical rate with an MWPM decoder (green curve) is minimized when these error rates are appropriately balanced. The logical error rate is calculated from the best fit of Eq. (4). Error bars are obtained by bootstrapping ($N = 10,000$ runs). Inset: Logical error rate per unit time, instead of per cycle.
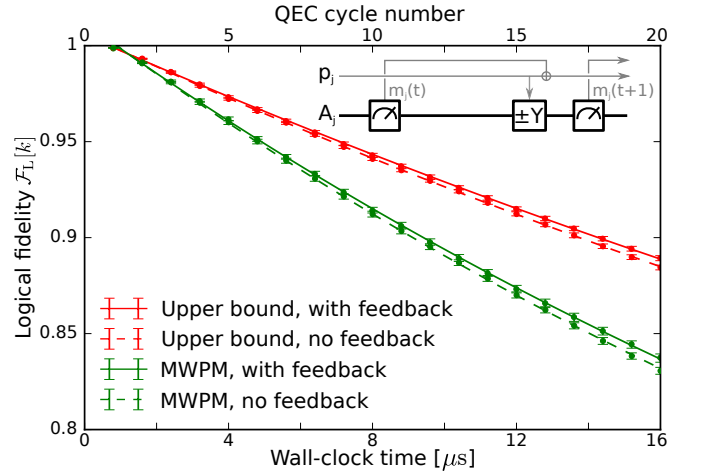

FIG. 3. Logical fidelity of Surface-17 with (solid) and without (dashed) an additional feedback scheme. The performance of a MWPM decoder (green) is compared to the decoder upper bound (red). Curves are compensated exponential fits to the data, and error bars are given by bootstrapping, with each point averaged over 10,000 runs. Inset: Method for implementing the feedback scheme. For each ancilla qubit A_j, we store a parity bit p_j, which decides the sign of the $R_y(\pi/2)$ rotation at the end of each coherent step. The time A_j spends in the ground state is maximized when p_j is updated each cycle t by XORing with the measurement result from cycle $t-1$, after the rotation of cycle t has been performed.

surement errors produced by control electronics. Thus, a path to reducing ϵ_L may be to decrease τ_{cycle}. Currently, the cycle is dominated by $\tau_m + \tau_d$. At fixed readout power, reducing τ_m and τ_d will reduce τ_{cycle} at the cost of increased readout infidelity ϵ_{RO} (described in Sec. IV B 6). We explore this trade-off in Fig. 2, using a linear-dispersive readout model [25], keeping $\tau_m = \tau_d$ and assuming no leftover photons. Because of the latter, $\epsilon_L^{(MWPM)}$ reduces from $1.07\%_c$ (Fig. 1) to $0.62\%_c$ at $\tau_m = 300$ ns. The minimum $\epsilon_L^{(MWPM)} = 0.55\%_c$ is achieved at around $\tau_m = 260$ ns. This is perhaps counterintuitive, as ϵ_{phys} reduces only $0.13\%_c$ while ϵ_{RO} increases 0.5%. However, it reflects the different sensitivity of the code to different types of errors. Indeed, $\epsilon_L^{(MWPM)}$ is smaller for $\tau_m = 200$ ns than for $\tau_m = 300$ ns, even though ϵ_{RO} increases to 5%. It is interesting to note that the optimal τ_m for quantum memory, which minimizes logical error per unit time, rather than per cycle, is $\tau_m = 280$ ns (Fig. 2 inset). This shows that different cycle parameters might be optimal for computation and memory applications.

Next, we consider the possibility to reduce ϵ_L using feedback control. Since T_1 only affects qubits in the excited state, the error rate of ancillas in Surface-17 is roughly two times higher when in the excited state. Thus, we consider using feedback to hold each ancilla in the ground state as much as possible. We do not consider feedback on data qubits, as the highly entangled logical states are equally susceptible to T_1.

The ancilla feedback can be achieved by replacing the $R_y(\pi/2)$ with $R_y(\pm\pi/2)$, depending on the value of a control bit. The optimal scheme (Fig. 3) is simple for current parameters, as feedback on one ancilla does not require input from other ancillas. [The qubit could be held for longer in the ground state if $R_y(\pi)$ could be applied immediately following measurement, but this is not possible due to leftover photons in the resonator and feedback latency [26].] Figure 3 shows the effect of this feedback using both the MWPM decoder and the decoder upper bound. We observe ϵ_L improve only $0.05\%_c$ in both cases. Future experiments might opt not to pursue these small gains in view of the technical challenges added by feedback control.

C. Projected improvement with advances in quantum hardware

We now estimate the performance increase that may result from improving the transmon relaxation and dephasing times via materials and filtering improvements. To model this, we return to $\tau_{cycle} = 800$ ns, and adjust T_1 values with both $T_\phi = 2T_1$ (common in experiment) and $T_\phi = \infty$ (all white-noise dephasing eliminated). We retain the same rates for coherent errors, readout infidelity, and photon-induced dephasing as in Fig. 1. Figure 4 shows the extracted ϵ_L and ϵ_{phys} over

the T_1 range covered. For the MWPM decoder (upper bound) and $T_\phi = 2T_1$, the memory figure of merit $\gamma_m = \epsilon_{\mathrm{phys}}/\epsilon_L$ increases from 1.3 (2) at $T_1 = 30\ \mu$s to 2 (5) at 100 μs. Completely eliminating white-noise dephasing will increase γ_m by 10% with MWPM and 30% at the upper bound.

A key question for any QEC code is how ϵ_L scales with code distance d. Computing power limitations preclude similar full density-matrix simulations of the $d = 5$ surface code Surface-49. However, we can approximate the error rate by summing up all lowest-order error chains (as calculated for the MWPM decoder), and deciding individually whether or not these would be corrected by a MWPM decoder (see App. E for details). Figure 5 shows the lowest-order approximation to the logical error rates of Surface-17 and -49 over a range of $T_1 = T_\phi/2$. Comparing the Surface-17 lowest-order approximation to the full quantumsim result shows good agreement and validates the approximation. We observe a lower ϵ_L for Surface-49 than for -17, indicating quantum fault tolerance over the T_1 range covered. The fault-tolerance figure of merit defined in [9], $\Lambda_t = \epsilon_L^{(17)}/\epsilon_L^{(49)}$, increases from 2 to 4 as T_1 grows from 30 to 100 μs.

As a rough metric of computational performance, we offer to compare ϵ_L (per cycle) to the error accrued by a physical qubit idling over $\tau_{\mathrm{g,1Q}}$. We define a metric for computation performance, $\gamma_c = (\epsilon_{\mathrm{phys}}\tau_{\mathrm{g,1Q}})/(\epsilon_L\tau_{\mathrm{cycle}})$ and $\gamma_c = 1$ as a computational break-even point. Clearly, using the QEC cycle parameters of Table I and even with T_1 improvements, neither Surface-17 nor -49 can breakeven computationally. However, including the readout acceleration recently demonstrated in [22], which allows $\tau_m = \tau_d = 100$ ns and $\tau_{\mathrm{cycle}} = 400$ ns, Surface-49 can cross $\gamma_c = 1$ by $T_1 = 40\ \mu$s. In view of first reports of T_1 up to 80 μs emerging for planar transmons [27, 28], this important milestone may be within grasp.

III. DISCUSSION

A. Computational figure of merit

We note that our metric of computational power is not rigorous, due to the different gate sets available to physical and logical qubits. Logical qubits can execute multiple logical X and Z gates within one QEC cycle, but require a few cycles for two-qubit and Hadamard gates (using the proposals of [12, 17]), and state distillation over many cycles to perform non-Clifford gates. As such, this metric is merely a rough benchmark for computational competitiveness of the QEC code. However, given the amount by which all distance-3 logical fidelities fall above this metric, we find it unlikely that these codes will outperform a physical qubit by any fair comparison in the near future.

B. Decoder performance

A practical question facing fault-tolerant quantum computing is how best to balance the trade-off between decoder complexity and performance. Past proposals for surface-code computation via lattice surgery [17] require the decoder to provide an up-to-date estimate of the Pauli error on physical qubits during each logical T gate. Because tracking Pauli errors through a non-Clifford gate is inefficient, however implemented, equivalent requirements will hold for any QEC code [29]. A decoder is thus required to process ancilla measurements from one cycle within the next (on average). This presents a considerable challenge for transmon-cQED implementations, as $\tau_{\mathrm{cycle}} < 1\ \mu$s. This short time makes the use of computationally intensive decoding schemes difficult, even if they provide lower ϵ_L.
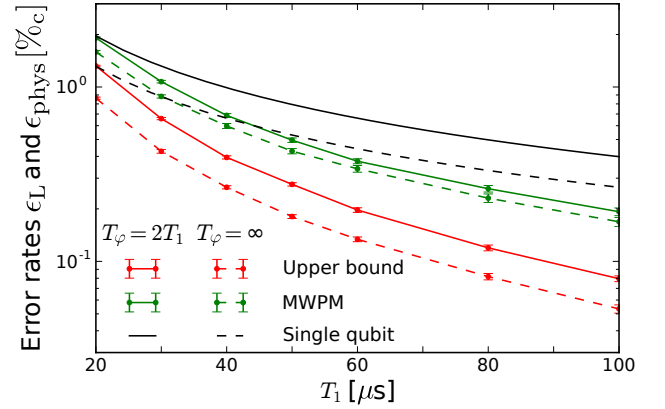

FIG. 4. T_1 dependence of the Surface-17 logical error rate (MWPM and UB) and the physical error rate. We either fix $T_\phi = 2T_1$ (solid) or $T_\phi = \infty$ (dashed). Logical error rates are extracted from a best fit of Eq. (4) to $\mathcal{F}_L[k]$ over $k = 1, \ldots, 20$ QEC cycles, averaged over $N = 50,000$ runs. Error bars are calculated by bootstrapping.

The leading strategy for decoding the surface code is MWPM using the blossom algorithm of Edmonds [10, 14, 30]. Although this algorithm is challenging to implement, it scales linearly in code distance [30]. The algorithm requires a set of weights (representing the probability that two given error signals are connected by a chain of errors) as input. An important practical question (see App. C) is whether these weights can be calculated on the fly, or must be precalculated and stored. On-the-fly weight calculation is more flexible. For example, it can take into account the difference in error rates between an ancilla measured in the ground and in the excited state. The main weakness of MWPM is the inability to explicitly detect Y errors. In fact, App. C shows that MWPM is nearly perfect in the absence of Y errors. The decoder efficiency η_d may significantly increase by extending MWPM to account for correlations between detected X and Z errors originating from Y errors [31, 32].

If computational limitations preclude a MWPM de-

coder from keeping up with τ_{cycle}, the look-up table decoder may provide a straightforward solution for Surface-17. However, at current physical performance, the η_d reduction will make Surface-17 barely miss memory break-even (Fig. 1). Furthermore, memory requirements make look-up table decoding already impractical for Surface-49. Evidently, real-time algorithmic decoding by MWPM or improved variants is an important research direction already at low code distance.

C. Other observations

The simulation results allow some further observations. Although we have focused on superconducting qubits, we surmise that the following statements are fairly general.

We observe that small quasi-static qubit errors are suppressed by the repeated measurement. In our simulations, the $1/f$ flux noise producing 0.01 radians of phase error per flux pulse on a qubit has a diamond norm approximately equal to the T_1 noise, but a trace distance 100 times smaller. As the flux noise increases ϵ_{L} by only 0.01 %$_c$, it appears ϵ_{L} is dependent on the trace distance rather than the diamond norm of the underlying noise components. Quasi-static qubit errors can then be easily suppressed, but will also easily poison an experiment if unchecked.

We further observe that above a certain value, ancilla and measurement errors have a diminished effect on ϵ_{L}. In our error model, the leading sources of error for a distance d code are chains of $(d-1)/2$ data qubit errors plus either a single ancilla qubit error or readout error, which together present the same syndrome as a chain of $(d+1)/2$ data qubit errors. An optimal decoder decides which of these chains is more likely, at which point the less-likely chain will be wrongly corrected, completing a logical error. This implies that if readout infidelity (ϵ_{RO}) or the ancilla error rate (ϵ_{anc}) is below the data qubit (ϵ_{phys}) error rate, $\epsilon_{\text{L}} \propto (\epsilon_{\text{anc}} + \epsilon_{\text{RO}})\epsilon_{\text{phys}}^{(d-1)/2}$. However, if ϵ_{RO} (ϵ_{anc}) > ϵ_{phys}, ϵ_{L} becomes independent of ϵ_{RO} (ϵ_{anc}), to lowest order. This can be seen in Fig. 2, where the error rate is almost constant as ϵ_{RO} exponentially increases. This approximation breaks down with large enough ϵ_{anc} and ϵ_{RO}, but presents a counterintuitive point for experimental design; ϵ_{L} becomes less sensitive to measurement and ancilla errors as these error get worse.

A final, interesting point for future surface-code computation is shown in Fig. 2: the optimal cycle parameters for logical error rates per cycle and per unit time are not the same. This implies that logical qubits functioning as a quantum memory should be treated differently to those being used for computation. This idea can be extended further: at any point in time, a large quantum computer performing a computation will have a set S_m of memory qubits which are storing part of a large entangled state, whilst a set S_c of computation qubits containing the rest of the state undergo operations. To minimize the probability of a logical error occurring on qubits within both S_c
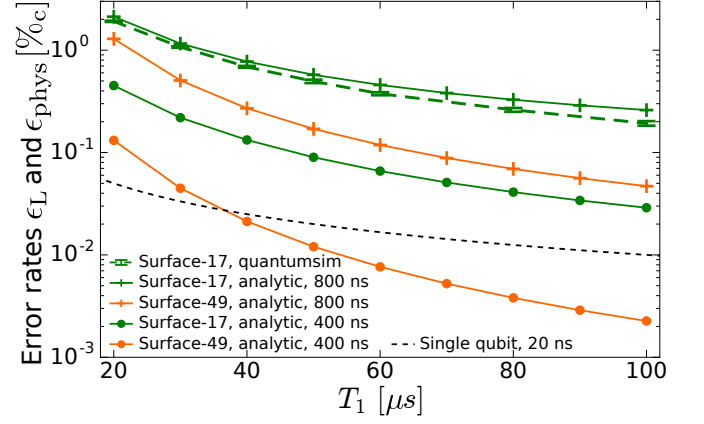

FIG. 5. Analytic approximation of ϵ_{L} for Surface-17 (green) and Surface-49 (orange) using a MWPM decoder. All plots assume $T_\phi = 2T_1$, and $\tau_{\text{cycle}} = 800$ ns (crosses) or 400 ns (dots). Numerical results for Surface-17 with $\tau_{\text{cycle}} = 800$ ns are also plotted for comparison (blue, dashed). Two physical-qubit computation metrics are given; the error incurred by a single qubit over the resting time of a single-qubit gate (black) and of a two-qubit gate (gray).

and S_m, the cycle time of the qubits in S_c can be reduced to minimize the rest time of qubits in S_m. As a simple example, consider a single computational qubit q_c and a single memory qubit q_m sharing entanglement. Operating all qubits at $\tau_{\text{cycle}} = 720$ ns to minimize ϵ_{L} would lead to a 1.09% error rate for the two qubits combined. However, shortening the τ_{cycle} of q_c reduces the time over which q_m decays. If q_c operates at $\tau_{\text{cycle}} = 600$ ns, the average error per computational cycle drops to 1.06%, as q_m completes only 5 cycles for every 6 on q_c. Although this is only a meager improvement, one can imagine that when many more qubits are resting than performing computation, the relative gain will be quite significant.

D. Effects not taken into account

Although we have attempted to be thorough in the detailing of the circuit, we have neglected certain effects. We have used a simple model for C-Z gate errors as we lack a full one obtained from two-qubit gate-set tomography [33]. Most importantly, we have neglected leakage, where a transmon is excited out of the two lowest energy states, i.e., out of the computational subspace. Previous experiments have reduced the leakage probability per C-Z gate to $\sim 0.3\%$ [34], and per single-qubit gate to $\sim 0.001\%$ [35]. Schemes have also been developed to reduce the accumulation of leakage [36]. Extending quantumsim to include and investigate leakage is a next target. Future simulations will also investigate the effect of spread in qubit parameters, both in space (i.e., variation of physical error rates between qubits) and time (e.g., T_1 fluctuations), and cross-talk effects such as residual couplings between nearest and next-nearest neighbor

transmons, qubit cross-driving, and qubit dephasing by measurement pulses targeting other qubits.

IV. METHODS

A. Simulated experimental procedure

1. Surface-17 basics

A QEC code can be defined by listing the data qubits and the stabilizer measurements that are repeatedly performed upon them [37]. In this way, Surface-17 is defined by a 3×3 grid of data qubits $\{D_0, \ldots D_8\}$. In order to stabilize a single logical qubit, $9 - 1 = 8$ commuting measurements are performed. The stabilizers are the weight-two and weight-four X- and Z-type parity operators $X_2 X_1$, $Z_3 Z_0$, $X_4 X_3 X_1 X_0$, $Z_5 Z_4 Z_2 Z_1$, $Z_7 Z_6 Z_4 Z_3$, $X_8 X_7 X_5 X_4$, $Z_8 Z_5$, and $X_7 X_6$, where X_j (Z_j) denotes the X (Z) Pauli operator acting on data qubit D_j. Their measurement is realized indirectly using nearest-neighbor interactions between data and ancilla qubits arranged in a square lattices, followed by ancilla measurements [Fig. 6(a)]. This leads to a total of 17 physical qubits when a separate ancilla is used for each individual measurement. We follow the circuit realization of this code described in [18], for which we give a schematic description in Fig. 6(b) (see App. A for a full circuit diagram).

In an experimental realization of this circuit, qubits will regularly accumulate errors. These errors cause the system state to evolve within the 512-dimensional Hilbert space of the data qubits. Multiple errors that occur within a short period of time (e.g., one cycle) form error 'chains' that spread across the surface. Errors on single qubits, or correlated errors within a small subregion of Surface-17, fail to commute with the stabilizer measurements, creating error signals that allow diagnosis and correction of the error via a decoder. However, errors that spread across more than half the surface in a short enough period of time are misdiagnosed, causing an error on the logical qubit when wrongly corrected [10]. The rate at which these logical errors arise is the main focus of this paper.

2. Protocol for measurement of logical error rates

As the performance measure of Surface-17, we study the fidelity of the logical qubit as a quantum memory. We describe our protocol with an example 'run' in Fig. 6. We initialize all qubits in $|0\rangle$ and perform $k = 1, 2, \ldots, 20$ QEC cycles [Fig. 6(b)]. Although this initial state is not a stabilizer eigenstate, the first QEC cycle projects the system into one of the 16 overlapping eigenstates within the $+1$ eigenspace for Z stabilizers, which form the logical $|0\rangle$ state [10]. This implies that, in the absence of errors, the first measurement of the Z stabilizers will be $+1$, whilst

that of the X stabilizers will be random. In the following cycles, ancilla measurements of each run [Fig. 6(c)] are processed using a classical decoding algorithm. The decoder computes a Pauli update after each QEC cycle [Fig. 6(d)]. This is a best estimate of the Pauli operators that must be applied to the data qubits to transform the logical qubit back to the logical $|0\rangle$ state. The run ends with a final measurement of all data qubits in the computational basis. From this 9-bit outcome, a logical measurement result is declared [Fig. 6(e)]. First, the four Z-type parities are calculated from the 9 data-qubit measurement outcomes and presented to the decoder as a final set of parity measurements. This ensures that the final computed Pauli update will transform the measurement results into a set that measures $+1$ for all Z stabilizers. This results in one of 32 final measurements, from which the value of a logical Z operator can be calculated to give the measurement result (any choice of logical operator gives the same result). The logical fidelity $\mathcal{F}_L[k]$ after k QEC cycles is defined as the probability of this declared result matching the initial $+1$ state.

At long times and with low error rates, Surface codes have a constant logical error rate ϵ_L. The fidelity $\mathcal{F}_L[k]$ is obtained by counting the probability of an odd number of errors having occurred in total (as two σ_x errors cancel) [20, 38]:

$$
\begin{aligned}
\mathcal{F}_L[k] &= 1 - \sum_{l \text{ even}} \binom{k}{l} \epsilon_L^l (1 - \epsilon_L)^{k-l} \\
&= 1 - \frac{1}{2} \sum_l \binom{k}{l} \epsilon_L^l (1 - \epsilon_L)^{k-l} (1 + (-1)^l) \\
&= \frac{1}{2}[1 + (1 - 2\epsilon_L)^k]. \quad (3)
\end{aligned}
$$

However, at small k, the decay is dominated by the majority vote, for which $\epsilon_L \propto (k\epsilon_{\text{phys}})^{(d+1)/2}$. For example, for all the Surface-17 decay curves, we observe a quadratic error rate at small k, as opposed to the linear slope predicted by (3). In order to correct for this, we shift the above equation in k, giving the compensated exponential decay fit function

$$
\mathcal{F}_L[k] = \frac{1}{2}[1 + (1 - 2\epsilon_L)^{k-k_0}]. \quad (4)
$$

This function fits well to data with $k \geq 3$ in all plots, and thus allows accurate determination of ϵ_L.

3. The quantumsim simulation package

Quantumsim performs calculations on full density matrices utilizing a graphics processing unit in a standard desktop computer. The simulation is set up as follows: the full density matrix of the nine data qubits is allocated in memory with all qubits initialized to $|0\rangle$. One- and two-qubit gates are applied to the density matrix as completely positive, trace preserving maps represented

by Pauli transfer matrices. When a gate involving an ancilla qubit must be performed, the density matrix of the system is dynamically enlarged to include that one ancilla.

Qubit measurements are simulated as projective and following the Born rule, with projection probabilities given by the squared overlap of the input state with the measurement basis states. In order to capture empirical measurement errors, we implement a black-box measurement model (Sec. IV B 6) by sandwiching the measurement between idling processes. The measurement projects the system to a product state of the ancilla and the projected sub-block of the full density matrix. We can therefore remove the ancilla from the density matrix and only store its state right after projection, and continue the calculation with the partial density matrix of the other qubits. Making use of the specific arrangement of the interactions between ancillas and data qubits in Surface-17, it is possible to apply all operations to the density matrix in such an order (shown in App. A) that the total size of the density matrix never exceeds $2^{10} \times 2^{10}$ (nine data qubits plus one ancilla), which allows relatively fast simulation. On our hardware, simulating one QEC cycle of Surface-17 with quantumsim takes 25 ms.

We highlight an important advantage of doing full density-matrix calculations with quantumsim. We do not perform projective measurements of the data qubits. Instead, after each cycle, we extract the diagonal of the data-qubit density matrix, which represents the probability distribution if a final measurement were performed. We leave the density matrix undisturbed and continue simulation up to $k = 20$. This is a very useful property of the full density-matrix approach, because having a probability distribution of all final readout events greatly reduces sampling noise.

Our measurement model includes a declaration error probability (see Sec. IV B 6), where the projected state of the ancilla after measurement is not the state reported to the decoder. Before decoding, we thus apply errors to the outcomes of the ancilla projections, and smear the probability distribution of the data qubit measurement. To then determine the fidelity averaged over this probability distribution, we present all 16 possible final Z-type parities to the decoder. This results in 16 different final Pauli updates, allowing us to determine correctness of the decoder for all 512 possible measurement outcomes. These are then averaged over the simulated probability distribution. This produces good results after about $\sim 10^4$ simulated runs.

A second highlight of quantumsim is the possibility to quantify the sub-optimality of the decoder. The fidelity of the logical qubit obtained in these numerical simulations is a combination of the error rates of the physical qubits and the approximations made by the decoder. Full density-matrix simulations make it possible to disentangle these two contributions. Namely, the fidelity is obtained by assigning correctness to each of the 512 possible readouts according to 16 outputs of the decoder, and

summing the corresponding probabilities accordingly. If the probabilities are known, it is easy to determine the 16 results that a decoder should output in order to maximize fidelity (i.e., the output of the best-possible decoder). This allows placing a decoder upper bound $\mathcal{F}_L^{max}$ on logical fidelity as limited by the physical qubits independent of the decoder. Conversely, it also allows quantifying sub-optimality in the decoder used. In fact, we can make the following reverse statement: if our measurement model did not include a declaration error, then we could use the simulation to find the final density matrix of the system conditioned on a syndrome measurement. From this, the simulation could output exactly the 16 results that give $\mathcal{F}_L^{max}$, so that quantumsim could thus be used as a maximum-likelihood decoder. In this situation, $\mathcal{F}_L^{max}$ would not only be an upper bound, but indeed the performance of the best-possible decoder. However, as we add the declaration errors after simulation, we can only refer to $\mathcal{F}_L^{max}$ as the decoder upper bound.

B. Error models

We now describe the error model used by quantumsim. App. B provides the parameters we used to simulate these processes.

1. Idling qubits

While idling for a time τ, a transmon in $|1\rangle$ can relax to $|0\rangle$. Furthermore, a transmon in superposition can acquire random quantum phase shifts between $|0\rangle$ and $|1\rangle$ due to $1/f$ noise sources (e.g., flux noise) and broadband ones (e.g., photon shot noise [39] and quasiparticle tunneling [40]). These combined effects can be parametrized by probabilities $p_1 = \exp(-\tau/T_1)$ for relaxation, and $p_\phi = \exp(-\tau/T_\phi)$ for pure dephasing. The combined effects of relaxation and pure dephasing lead to decay of the off-diagonal elements of the qubit density matrix. We model dephasing from broadband sources in this way, taking for T_ϕ the value extracted from the decay time T_2 of standard echo experiments:

$$\frac{1}{T_2} = \frac{1}{T_\phi} + \frac{1}{2T_1}. \tag{5}$$

We model $1/f$ sources differently, as discussed below.

2. Dephasing from photon noise

The dominant broadband dephasing source is the shot noise due to photons in the readout resonator. This dephasing is present whenever the coupled qubit is brought into superposition before the readout resonator has returned to the vacuum state following the last measurement. This leads to an additional, time-dependent pure dephasing (rates given in App. B 2).
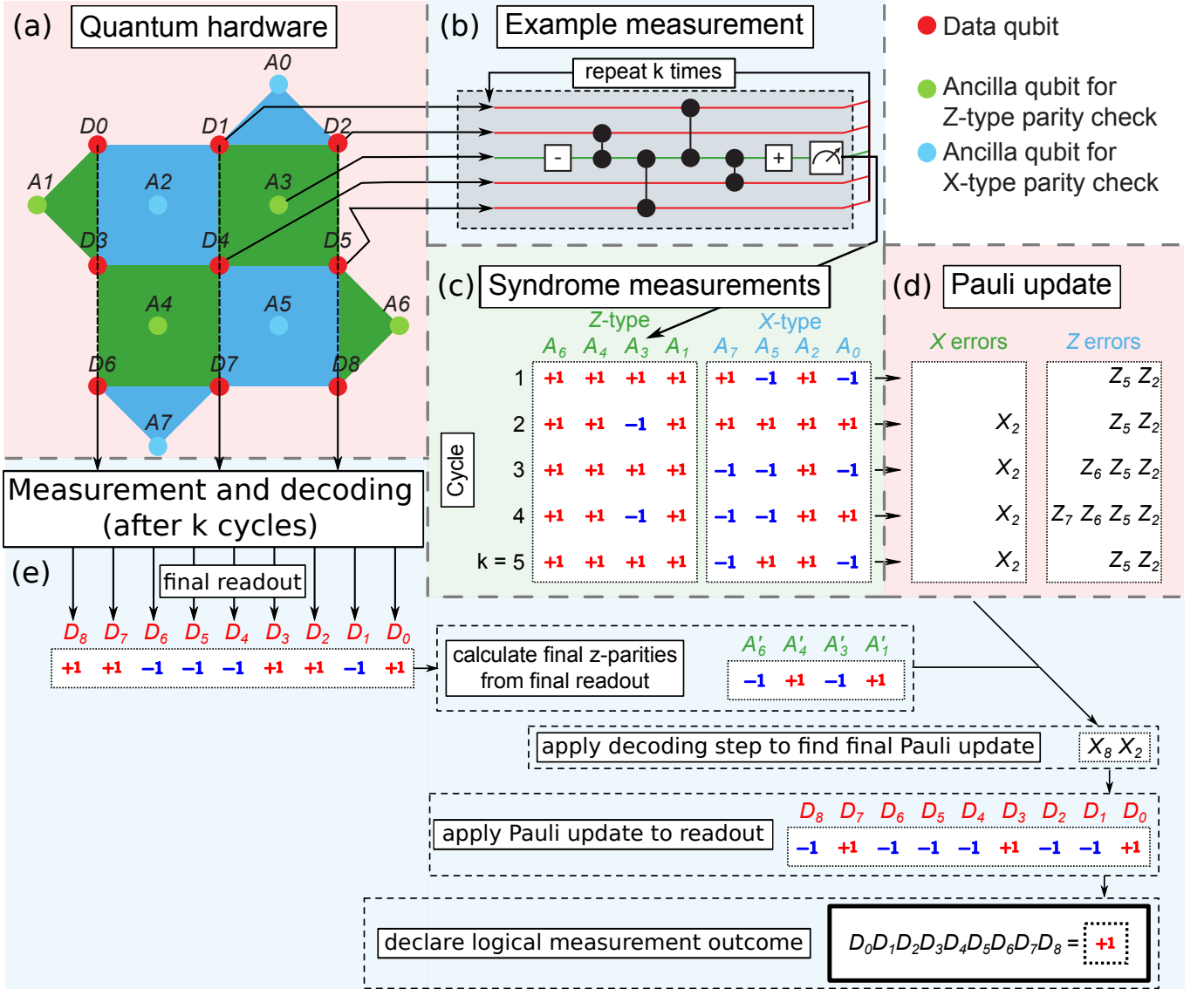
FIG. 6. Schematic overview of the simulated experiment. (a) 17 qubits are arranged in a surface code layout (legend top-right). The red data qubits are initialized in the ground state $|0\rangle$, and projected into an eigenstate of the measured X- (blue) and Z- (green) type stabilizer operators. (b) A section of the quantum circuit depicting the four-bit parity measurement implemented by the A_3 ancilla qubit ($+/-$ refer to $R_y(\pm\pi/2)$ single-qubit rotations). The ancilla qubit (green line, middle) is entangled with the four data qubits (red lines) to measure $Z_1 Z_2 Z_4 Z_5$. Ancillas are not reset between cycles. Instead, the implementation relies on the quantum non-demolition nature of measurements. The stabilizer is then the product of the ancilla measurement results of successive cycles. This circuit is performed for all ancillas and repeated k times before a final measurement of all (data and ancilla) qubits. (c) All syndrome measurements of the k cycles are processed by the decoder. (d) After each cycle, the decoder updates its internal state to represent the most likely set of errors that occurred. (e) After the final measurement, the decoder uses the readout from the data qubits, along with previous syndrome measurements, to declare a final logical state. To this end, the decoder processes the Z-stabilizers obtained directly from the data qubits, finalizing its prediction of most likely errors. The logical parity is then determined as the product of all data qubit parities ($\prod_{j=0}^{8} D_j$) once the declared errors are corrected. The logical fidelity $\mathcal{F}_\mathrm{L}$ is the probability that this declaration is the same as the initial state ($|0\rangle$).

3. One-qubit Y rotations

We model y-axis rotations as instantaneous rotations sandwiched by idling periods of duration $\tau_{\mathrm{g,1Q}}/2$. The errors in the instantaneous gates are modeled from process matrices measured by gate-set tomography [33, 41]

in a recent experiment [20]. We assume that these errors are Markovian.

4. Dephasing of flux-pulsed qubits

During the coherent step, transmons are repeatedly moved in frequency away from their sweetspot using flux pulses, either to implement a C-Z gate or to avoid one. Away from the sweetspot, transmons become first-order sensitive to flux noise, which causes an additional random phase shift. As this noise typically has a $1/f$ power spectrum, the largest contribution comes from low-frequency components that are essentially static for a single run, but fluctuating between different runs. In our simulation, we approximate the effect of this noise through ensemble averaging, with quasi-static phase error added to a transmon whenever it is flux pulsed. Gaussian phase errors with the variance calculated in App. B 4 are drawn independently for each qubit and for each run.

5. C-Z gate error

The C-Z gate is achieved by flux pulsing a transmon into the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing with another, where the 2 denotes the second-excited state of the fluxed transmon. Holding the transmons here for $\tau_{\text{g},2\text{Q}}$ causes the probability amplitudes of $|01\rangle$ and $|11\rangle$ to acquire phases [42]. Careful tuning allows the phase ϕ_{01} acquired by $|01\rangle$ (the single-qubit phase $\phi_{1\text{Q}}$) to be an even multiple of 2π, and the phase ϕ_{11} acquired by $|11\rangle$ to be π extra. This extra phase acquired by $|11\rangle$ is the two-qubit phase $\phi_{2\text{Q}}$. Single- and two-qubit phases are affected by flux noise because the qubit is first-order sensitive during the gate. Previously, we discussed the single-qubit phase error. In App. B 5, we calculate the corresponding two-qubit phase error $\delta\phi_{2\text{Q}}$. Our full (but simplistic) model of the C-Z gate consists of an instantaneous C-Z gate with single-qubit phase error $\delta\phi_{1\text{Q}}$ and two-qubit phase error $\delta\phi_{2\text{Q}} = \delta\phi_{1\text{Q}}/2$, sandwiched by idling intervals of duration $\tau_{\text{g},2\text{Q}}/2$.

6. Measurement

We model qubit measurement with a black-box description using parameters obtained from experiment. This description consists of the eight probabilities for transitions from an input state $|i\rangle \in \{|0\rangle, |1\rangle\}$ into pairs $(m, |o\rangle)$ of measurement outcome $m \in \{+1, -1\}$ and final state $|o\rangle \in \{|0\rangle, |1\rangle\}$. By final state we mean the qubit state following the photon-depletion period. Input superposition states in the computational bases are first projected to $|0\rangle$ and $|1\rangle$ following the Born rule. The probability tree (the butterfly) is then used to obtain an output pair $(m, |o\rangle)$. The six free parameters in this black box model can be mapped to a six-parameter model consisting of decay periods before and after a point at which the qubit is perfectly projected, as well as probabilities for wrongly declaring the result of the projective measurement. App. B 6 describes a scheme for measuring

these butterfly parameters and mapping them to the six-parameter model.

Appendix A: Full circuit diagram for Surface-17 implementation

The quantum circuit [18] (Fig. 7) consists of $R_y(\pi/2)$ ("+") and $R_y(-\pi/2)$ ("−") rotations, C-Z gates, and ancilla measurements. The coherent steps of the X and Z ancillas are pipelined (shifted in time with respect to each other) to prevent transmon-transmon avoided crossings. As long as $\tau_{\text{m}} + \tau_{\text{d}} \geq \tau_{\text{c}}$, no time is lost due to this separation.

In a simulation of the given circuit, gates on different qubits commute and may be applied to the density matrix in any order, regardless of the times at which they are performed in an experiment. As described in Sec. IV A 3, by simulating gates in a specific order (Fig. 8), one can ensure that only one ancilla is ancilla is entangled with the data qubits at any point in the simulation. This allows a reduction in the maximum size of the density matrix from $2^{17} \times 2^{17}$ to $2^{10} \times 2^{10}$.

Appendix B: Parameters of error models

This appendix provides mathematical details of the sources of error described in the main text. Standard values for the parameters used throughout the text are given in Table II.

In the quantumsim module, all gates are applied in the Pauli transfer matrix representation [45]. These are given in the form

$$(R_\Lambda)_{ij} = \frac{1}{2}\text{Tr}\left(\sigma_i \Lambda \sigma_j\right), \tag{B1}$$
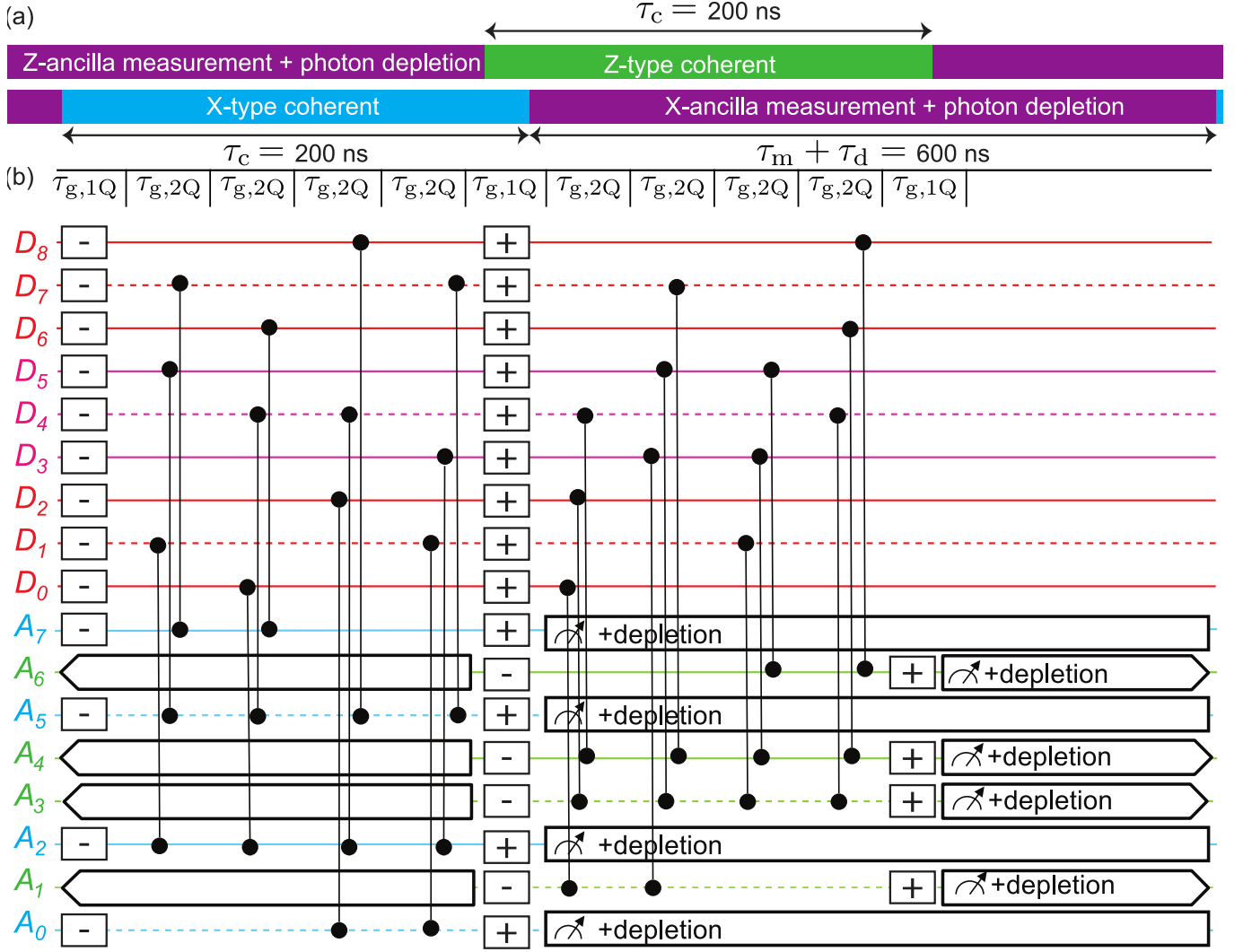
FIG. 7. The quantum circuit for Surface-17 syndrome measurement used in all simulations. (a) Outline of the timing of the standard circuit, including the time shift between X- and Z-type stabilizer measurements described by [18]. Qubit labels correspond to the position in Fig. 6. (b) Full quantum circuit of the QEC cycle. The C-Z gates within each group are slightly offset horizontally for visibility (in reality they are performed simultaneously).

Parameter	Symbol	Value	Reference
In-axis rotation error	p_{axis}	10^{-4}	[20]
In-plane rotation error	p_{plane}	5×10^{-4}	[20]
$1/f$ flux noise.	A	$(1\mu\Phi_0)^2$	[43, 44]
Readout infidelity	ϵ_{RO}	5×10^{-3}	[19]
Photon relaxation time	$1/\kappa$	250 ns	[19]
Dispersive shift	χ/π	-2.6 MHz	[19]
photon # post-measurement	n_0	0.8 photons	[19]

TABLE II. Standard parameters of error models used in quantumsim, unless indicated otherwise.

where matrices σ_i are the Pauli operators: $\sigma_0 = I$, $\sigma_1 = X$, $\sigma_2 = Y$ and $\sigma_3 = Z$.

1. Qubit idling

Idling qubits are described by the amplitude-phase damping model [46], corresponding to the transfer matrices

$$R_{\Lambda_{T_1}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-p_1} & 0 & 0 \\ 0 & 0 & \sqrt{1-p_1} & 0 \\ p_1 & 0 & 0 & 1-p_1 \end{pmatrix} \quad \text{(B2)}$$

$$R_{\Lambda_{T_\phi}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-p_\phi} & 0 & 0 \\ 0 & 0 & \sqrt{1-p_\phi} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad \text{(B3)}$$
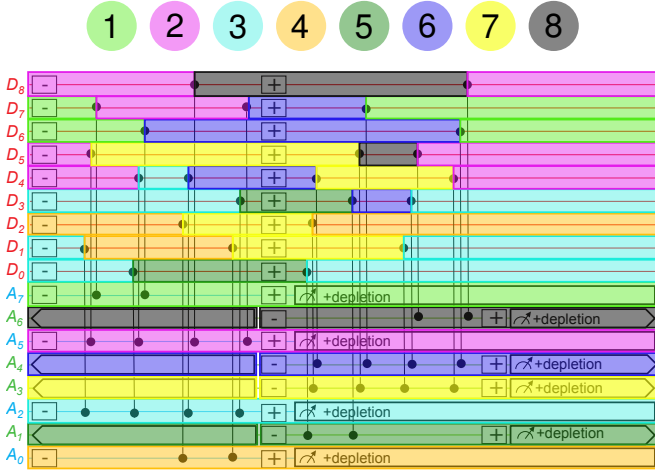
FIG. 8. Isolation of ancilla interactions in the Surface-17 circuit given in Fig. 7. Throughout a simulation, quantumsim stores the density matrix of all data qubits. Each error correction cycle is split up into 8 steps as labeled. In each step, a single ancilla qubit is added to the density matrix, the correspondingly colored pieces of the circuit are executed, and the ancilla is read out and removed from the density matrix. This scheme is only possible because on each data qubit all gates are executed in order. Note that steps after the final C-Z gate on a data qubit are executed during the next cycle.

Idling for a duration t is thus described by

$$R_{AP(t)} = R_{\Lambda_{T_1}} R_{\Lambda_{T_\phi}} \qquad (B4)$$

with $p_1 = 1 - e^{-t/T_1}$ and $p_\phi = 1 - e^{-t/T_\phi}$.

2. Photon decay

In the presence of photons in a readout resonator, the coupled qubit is affected according to the effective stochastic master equation [25]:

$$\frac{d\rho}{dt} = -i\frac{B}{2}[\sigma_z, \rho] + \frac{\Gamma_d}{2}\mathcal{D}[\sigma_z]\rho.$$

Here, ρ is the qubit density matrix, $\mathcal{D}[X]$ is the Lindblad operator $\mathcal{D}[X]\rho = X\rho X^\dagger - \frac{1}{2}X^\dagger X\rho - \frac{1}{2}\rho X^\dagger X$, $B = 2\chi\text{Re}(\alpha_g\alpha_e^*)$ is the measurement-induced detuning (Stark shift), and $\Gamma_d = 2\chi\text{Im}(\alpha_g\alpha_e^*)$ is the measurement-induced dephasing, with α_i the qubit-state-dependent photon field in the resonator and 2χ the qubit frequency shift per photon. At time $t - t_g$ after the qubit superposition is created,

$$\alpha_g\alpha_e^* = \alpha(t_m)\exp\left(-\kappa\left(t - t_m\right)\right)\exp\left(2i\chi\left(t - t_g\right)\right),$$

with $t - t_m$ the time since the end of measurement excitation pulse. Integrating over the interval $[t_1, t_2]$ gives a dephasing term with coefficient

$$
\begin{aligned}
p_{\phi,\text{photon}} &= \exp\left(-\int_{t_1}^{t_2}\Gamma_d(t)dt\right) \\
&= \exp\Bigg(2\chi\alpha(0)\exp(\kappa(t_m - t_g)) \\
&\quad \times \left[\frac{e^{-\kappa t}}{4\chi^2+\kappa^2}[-\kappa\sin(2\chi t) - 2\chi\cos(2\chi t)]\right]_{t_1-t_g}^{t_2-t_g}\Bigg).
\end{aligned}
$$

This dephasing is then implemented via the same Pauli transfer matrix as (B3).

3. Single-qubit $R_y(\pi/2)$ rotations

Single-qubit rotations are modeled by sandwiching an instantaneous Pauli transfer matrix, representing the rotation, with periods of duration $\tau_{g,1Q}/2$ of amplitude and phase damping. This allows to model the gate for different T_1 and T_ϕ. However, comparison of this model with Pauli transfer matrices obtained from gate-set tomography experiments shows that actual gates are more accurately described when adding a phenomenological depolarizing noise to the instantaneous part. In the Bloch sphere, this decay corresponds to shrinking toward the origin, with factor $1 - p_{\text{axis}}$ along the y axis and $1 - p_{\text{plane}}$ along the x- and z-axes. We thus model

$$R_{R_y(\pi/2)} = R_{AP(\tau_{g,1Q}/2)}R'_{R_y(\pi/2)}R_{\text{dep}}R_{AP(\tau_{g,1Q}/2)},$$

where

$$R_{\text{dep}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - p_{\text{plane}} & 0 & 0 \\ 0 & 0 & 1 - p_{\text{axis}} & 0 \\ 0 & 0 & 0 & 1 - p_{\text{plane}} \end{pmatrix},$$

and $R'_{R_y(\pi/2)}$ is the Pauli transfer matrix describing a perfect $\pi/2$ rotation around the y axis.

4. Flux noise

Shifting the transmon from its sweetspot $f_{q,\text{max}}$ to a lower frequency

$$f_q(t) = (f_{q,\text{max}} + E_C)\sqrt{|\cos(\pi\Phi(t)/\Phi_0)|} - E_C$$

makes it first-order sensitive to flux noise, with sensitivity

$$\frac{\partial f_q}{\partial \Phi} = \frac{-\pi}{2\Phi_0}(f_q + E_C)\tan\left(\frac{\pi\Phi}{\Phi_0}\right).$$

Here, Φ is the flux bias and $\Phi_0 = h/2e$ is the flux quantum. For a deviation of $\delta\Phi$, the pulsed transmon incurs a phase error

$$\delta\phi = -2\pi\tau_{g,2Q}\frac{\partial f_q}{\partial \Phi}\delta\Phi.$$

Flux noise has a characteristic (single-sided) spectral density

$$S_\Phi(f) \approx A/f,$$

where $A \approx (1 \ \mu\Phi_0)^2$ with f in Hz. We model this noise as quasi-static over the duration ($1/f_{\min} \sim 20 \ \mu$s, or 20 QEC cycles) of individual runs, but fluctuating between subsequent runs ($1/f_{\max} \sim 20$ sec, or 10^5 runs at 200 μs intervals). The root-mean-square (rms) fluctuations of flux are therefore

$$
\begin{aligned}
\delta\Phi_{\mathrm{rms}} &= \left(\int_{f_{\min}}^{f_{\max}} S_\Phi(f) \, df \right)^{1/2} \\
&= A(\ln(f_{\max}/f_{\min}))^{1/2} \\
&\approx 4 \ \mu\Phi_0.
\end{aligned}
$$

For our quantum circuit based on [18], we estimate the corresponding rms phase error induced in a pulsed transmon to be

$$\delta\phi_{\mathrm{rms}} \approx 0.01 \ \mathrm{rad}.$$

5. C-Z gates

We now focus on the two-qubit phase error. For an adiabatic gate,

$$\phi_{2Q} = \phi_{11} - \phi_{01} = -2\pi \int_{t_1}^{t_2} \zeta(t) dt,$$

with t_1 and $t_2 = t_1 + \tau_{\mathrm{g,2Q}}$ the start and end of the gate and ζ the time-dependent frequency deviation of the lower branch of the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing from the sum of frequencies for $|01\rangle$ and $|10\rangle$. Near the flux center Φ_{c} of the $|11\rangle - |02\rangle$ avoided crossing,

$$\zeta \approx \beta(\Phi - \Phi_{\mathrm{c}}) - \sqrt{\beta^2(\Phi - \Phi_{\mathrm{c}})^2 + (2J/2\pi)^2},$$

where $2J/2\pi \sim 50$ MHz is the minimum splitting between $|11\rangle$ and $|02\rangle$, and

$$\beta = \frac{1}{2}\frac{\partial f_{\mathrm{q}}}{\partial \Phi}|_{\Phi=\Phi_{\mathrm{c}}}.$$

Differentiating with respect to Φ at Φ_{c} gives

$$\frac{\partial \zeta}{\partial \Phi}|_{\Phi=\Phi_{\mathrm{c}}} = \beta.$$

To estimate the $\delta\phi_{2Q}$ error, we make the following simplification: we replace the exact trajectory created by the flux pulse by a shift to $\Phi = \Phi_{\mathrm{c}} + \delta\Phi$ with duration $\tau_{\mathrm{g,2Q}}$. For a deviation of $\delta\Phi$,

$$\delta\phi_{2Q} \approx -2\pi\tau_{\mathrm{g,2Q}}\frac{\partial \zeta}{\partial \Phi}|_{\Phi=\Phi_{\mathrm{c}}}\delta\Phi.$$

Note that this two-qubit phase error is correlated with the single-qubit phase error on the fluxed transmon. The former is smaller by a factor ≈ 2.

6. Measurement

The probabilities $\epsilon_i^{m,o}$ are calibrated using the statistics of outcomes in back-to-back measurements (a followed by b) with the qubit initialized in $|i\rangle$.

$$
\begin{aligned}
\mathrm{P}(m_{\mathrm{a}} = +1)_i &= \epsilon_i^{+1,0} + \epsilon_i^{+1,1}, \\
\mathrm{P}(m_{\mathrm{a}} = +1)_i &= \epsilon_i^{-1,0} + \epsilon_i^{-1,1}, \\
\mathrm{P}(m_{\mathrm{b}} = m_{\mathrm{a}} = +1)_i &= \left(\epsilon_0^{+1,0} + \epsilon_0^{+1,1}\right)\epsilon_i^{+1,0} \\
&\quad + \left(\epsilon_1^{+1,0} + \epsilon_1^{+1,1}\right)\epsilon_i^{+1,1}, \\
\mathrm{P}(m_{\mathrm{b}} = -m_{\mathrm{a}} = +1)_i &= \left(\epsilon_0^{+1,0} + \epsilon_0^{+1,1}\right)\epsilon_i^{-1,0} \\
&\quad + \left(\epsilon_1^{+1,0} + \epsilon_1^{+1,1}\right)\epsilon_i^{-1,1}, \\
\mathrm{P}(-m_{\mathrm{b}} = m_{\mathrm{a}} = +1)_i &= \left(\epsilon_0^{-1,0} + \epsilon_0^{-1,1}\right)\epsilon_i^{+1,0} \\
&\quad + \left(\epsilon_1^{-1,0} + \epsilon_1^{-1,1}\right)\epsilon_i^{+1,1}, \\
\mathrm{P}(-m_{\mathrm{b}} = -m_{\mathrm{a}} = +1)_i &= \left(\epsilon_0^{-1,0} + \epsilon_0^{-1,1}\right)\epsilon_i^{-1,0} \\
&\quad + \left(\epsilon_1^{-1,0} + \epsilon_1^{-1,1}\right)\epsilon_i^{-1,1}.
\end{aligned}
$$

We obtain the six free parameters of the black-box description from these 12 equations, using experimental values on the left-hand side [47]. Table III shows the values used, achieved in a recent experiment [19]. We represent

Probability	Value	Probability	Value
$\epsilon_0^{+1,0}$	0.9985	$\epsilon_1^{+1,0}$	0.0050
$\epsilon_0^{+1,1}$	0.0000	$\epsilon_1^{+1,1}$	0.0015
$\epsilon_0^{-1,0}$	0.0015	$\epsilon_1^{-1,0}$	0.0149
$\epsilon_0^{-1,1}$	0.000	$\epsilon_1^{-1,1}$	0.9786

TABLE III. Measurement butterfly matching a recent characteristic experiment [19] using a Josephson parametric amplifier [48] in phase-preserving mode as the front end of the readout amplification chain.

this black-box model in simulation by a perfect projective measurement sandwiched between two idling periods described by Pauli transfer matrices containing the measurement-induced excitation and de-excitation probabilities (four parameters). Additionally, the result of the perfect projection is wrongly declared to the decoder, with (potentially) different probabilities for +1 and −1 outcomes (two parameters). These parameters are chosen so that the model reproduces the butterfly probabilities.

We note that the excitation and de-excitation parameters following from Tab. III are very well explained by assuming unmodified amplitude-phase damping during the measurement period, together with an outcome-independent declaration error of 0.15%. We use this approximation to extrapolate measurement performance to different values of T_1 and T_ϕ.

Reduction of measurement time is expected to reduce assignment fidelity. For the results presented in Fig. 2, we assume a simplified model for measurement, following Ref. 25. A constant drive pulse of amplitude ϵ and

tuned to the bare resonator frequency, $\Delta_r = 0$, excites the readout resonator for time τ_{m}. The dynamics of the resonator is dependent on the transmon state (we approximate linear behavior), and the transmitted signal is amplified and detected in a homodyne measurement as a noisy transient. This transient is processed by a linear classifier, which declares the measurement outcome. For resonator depletion, we use a two-step clearing pulse with amplitude ϵ_{c1} and ϵ_{c2}, each active for $\tau_{\mathrm{d}}/2$ and chosen (by numerical minimization) so that, at the end of the depletion pulse, the transients for both transmon states return to zero. While the resonator dynamics is easily found if the transmon is in the ground state, amplitude damping of the transmon in the excited state leads to non-deterministic behavior. We thus numerically obtain an ensemble of noisy transients for each input qubit state, and optimize the decision boundary of the linear classifier for this ensemble. Generating a second verification ensemble, the "butterfly" of the measurement setup is estimated.

The dynamics of the resonator is determined by the resonator linewidth κ as well as the dispersive shift χ. We chose the parameters of the setup used in [19], $1/\kappa = 250\,\mathrm{ns}$ and $\chi/\pi = -2.6\,\mathrm{MHz}$. The signal-to-noise ratio of the detected transient is reduced by the quantum efficiency $\eta = 12.5\%$. The driving strength ϵ is chosen to approximate the "butterfly" used in most of the main text, and corresponds to a steady-state average photon population of about $\bar{n} = 15$. We then keep ϵ constant while changing the measurement time, keeping $\tau_{\mathrm{m}} = \tau_{\mathrm{d}}$, to obtain the butterflies used in the density matrix simulation. We ignore effects leading to measurement-induced mixing and non-linearity of the readout resonator. Finally, since these simulations do not allow to make a realistic prediction about residual photon numbers achievable in experiments, we ignore this effect when using these results.

Appendix C: Homemade MWPM decoder with asymmetric weight calculation

Every QEC code requires a decoder to track the most likely errors consistent with a given set of stabilizer measurements. The MWPM decoder has gained popularity since it was shown to have threshold values above 1% [14]. The motivation behind MWPM is that single X or Z errors on data qubits in the bulk of a surface-code fabric cause changes of two stabilizers in the code. These signals can then be considered vertices on a graph, with the error the edge connecting them. Errors in measurement, or errors on a single ancilla qubit, behave as changes in the stabilizer that are separated in time. Multiple errors that would join the same vertices create longer paths in the graph, of which an experiment only records the endpoints. Thus, the problem becomes that of finding the most likely set of generating errors given the error signals that mark their ends. This is made slightly simpler, as in

the surface code any chain of errors that forms a closed loop does not change the logical state. This implies that all paths that connect two points are equivalent, and can be considered together. The problem then is to join error signals, either in pairs, or to a 'boundary' vertex. The latter corresponds to errors on data qubits at the boundary, which belong to only one X or Z stabilizer. This pairing P should be chosen as the most likely combination of single-qubit errors that could generate the measured error signals. This has then been reduced to the problem of minimum-weight perfect matching on a graph, which can be solved in polynomial time by the blossom algorithm [10, 49].

The MWPM decoder we use differs from previous methods by its weight calculation. As part of the decoding process, it is required to calculate to some degree of accuracy [50] the probability p_{e_1,e_2} of two measured error signals e_1 and e_2 being connected by a chain of individual logical errors. This is then converted to a weight $w_{e_1,e_2} = -\log(p_{e_1,e_2})$, which form the input to the blossom algorithm of Edmonds to find the most likely matching of error signals [10, 49]. An exact calculation of p_{e_1,e_2} requires a sum over all such chains between e_1 and e_2 that do not cross the boundary (these are equivalent modulo stabilizer operators that do not change the logical state). In this appendix we detail a method of computing this sum, and approximations to make it viable within the runtime of the experiment.

Let us define the ancilla graph $\mathcal{G}_A = (V_A, E_A)$ containing a vertex $v \in V_A$ for every ancilla measurement, and an edge $e \in E_A$ connecting $v, u \in V_A$ if a single component (gate, single-qubit rest period, or faulty measurement) in the simulation can cause the u and v measurements to return an error. We include a special 'boundary' vertex v_B, to which we connect another vertex v if single components can cause errors on v alone. Then, to each edge e we associate a probability p_e, being the sum of the probabilities of each component causing this error signal. These error rates can be obtained directly from quantumsim, by cutting the circuit at each C-Z gate and measuring the decay of single qubits between. Then, for a given experiment with given syndrome measurements, let us define the syndrome graph $\mathcal{G}_S = (V_S, E_S)$ containing a vertex $v \in V_S$ for each syndrome measurement that records an error, and an edge $\lambda_{u,v} \in E_S$ connecting $u, v \in V_S$ if u and v are either both X ancilla qubits or both Z ancilla qubits. To each edge $\lambda_{u,v}$ we associate a probability $p_{u,v}$ given by the sum of the probabilities of a chain of errors causing error signals solely on u and v.

If we assume that single-qubit errors are uncorrelated, we have to lowest order

$$p_{u,v} \approx \sum_{\text{paths } (e_1,e_2,...,e_n) \text{ between } u \text{ and } v} \prod_{j=1}^{n} p_{e_j}, \quad \text{(C1)}$$

Let A_A be the adjacency matrix on $\mathcal{G}_A$ weighted by the probabilities p_e (i.e., $(A_A)_{u,v} = p_e$ with e connecting u and v), and A_S the same for $\mathcal{G}_S$. Then, the above be-

comes

$$A_S = A_A + A_A^2 + A_A^3 + \cdots = \frac{1}{1 - A_A} - 1, \quad \text{(C2)}$$

noting that A_S contains a subset of the indices that are used to construct A_A.

The boundary must be treated specially in the above calculation. For the purposes of the surface code, the boundary can be described as a single vertex which has no limit on the number of other vertices it may pair to [10]. For the purposes of weight calculation, any path that passes through the boundary is already counted by pairing both end vertices to the boundary. This can be treated by making $\mathcal{G}_A$ directed, and breaking the symmetry $A_A^T = A_A$. In particular, either $(A_A)_{v_B,u} = 0$ for all u or $(A_A)_{u,v_B} = 0$ for all u.

The above calculation requires inversion of a $N_{\text{mat}} \times N_{\text{mat}}$ matrix, with N_{mat} the total number of ancilla measurements per experiment. Furthermore, as ancilla error rates depend upon the previous ancilla state, elements in A_A are not completely known until the previous cycle. This implies that in an actual computation with run-time decoding, this inversion would need to be completed within a few microseconds (with a transmon-cQED architecture), which is practically unfeasible. We suggest two approximations that can be made to shorten the decoding time. The first is to average all errors over the ancilla population, ignoring any asymmetry in the system. The adjacency matrix is now the same for any experiment, and can be precalculated and stored as a look-up table for the run-time decoder. We call this the decoder with symmetrized weights. The size of such a look-up table scales poorly with the number of qubits and the number of cycles. However, $(A_S)_{u,v}$ is approximately invariant under simultaneous translation of u and v (excluding boundary effects). This implies that a precalculated A_S can be vastly compressed, making this method feasible.

The second approximation to the full A_S calculation is to perform it iteratively. We divide our graph $\mathcal{G}_A$ ($\mathcal{G}_S$) by time steps; let $\mathcal{G}_A^t$ ($\mathcal{G}_S^t$) be the subgraph of $\mathcal{G}_A$ ($\mathcal{G}_S$) containing only ancillas measured before time step t, and let $\partial\mathcal{G}_A^t$ ($\partial\mathcal{G}_S^t$) be the subgraph of $\mathcal{G}_A$ ($\mathcal{G}_S$) containing only ancillas measured during time step t. Then, if we assume we have an approximation to the matrix A_S^t (being the adjacency matrix of $\mathcal{G}_S^t$), we can approximate

$$A_S^{t+1} \approx \begin{pmatrix} A_S^t & C_S^{t+1} \\ \left(C_S^{t+1}\right)^T & \left(\mathbf{1} - \partial A_A^{t+1}\right)^{-1} \end{pmatrix} \quad \text{(C3)}$$

to lowest order in physical errors. Here, ∂A_A^{t+1} is the weighted adjacency matrix on $\partial\mathcal{G}_A^{t+1}$, and the coupling matrix C_S^{t+1} is approximated by

$$C_S^{t+1} = A_S^t C_A^{t+1} \left(\mathbf{1} - \partial A_A^{t+1}\right)^{-1}, \quad \text{(C4)}$$

with C_A^{t+1} the adjacency matrix containing only edges between $\partial\mathcal{G}_A^{t+1}$ and $\mathcal{G}_A^t$. This procedure corresponds to
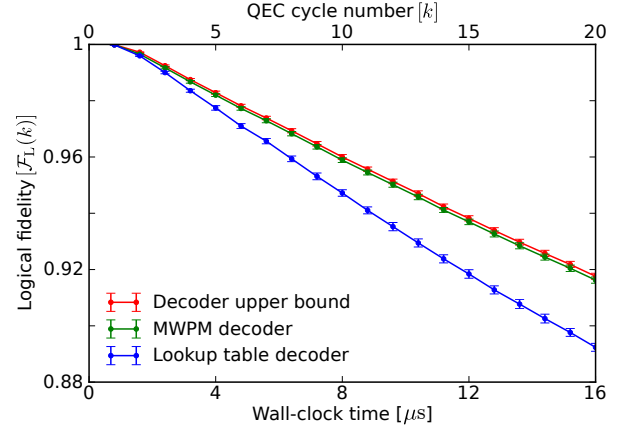

FIG. 9. Simulation of the experimental protocol used throughout the work, but using an error model that has Y errors and readout infidelity removed. With these errors absent, the MWPM decoder achieves the decoder upper bound within simulation error. The look-up table approach (blue) retains some inaccuracy beyond this.

a sum over all paths that are made by moving within $\partial\mathcal{G}_A^{t+1}$, shifting back in time to $\mathcal{G}_A^t$, and then taking any precalculated path in $\mathcal{G}_A^t$. C_A^{t+1} and $\left(\mathbf{1} - \partial A_A^{t+1}\right)^{-1}$ can be precalculated, and so the runtime computation requirement is reduced to the product in Eq. (C4). This in turn can be sparsified, as C_A^{t+1} only contains connections to vertices in $\mathcal{G}_A^t$ close to the time boundary, and we can delete all terms in A_S^t that do not connect from these vertices to errors.

We have used the second method for our MWPM decoder, as we expect the error from neglecting higher-order combinations of errors to be small. In order to check this assumption, in Fig. 9 we repeat our simulation protocol with a modified physical error model that excludes all Y and measurement errors. We see that in the absence of these errors, the MWPM decoder performs within the error margin of the decoder upper bound. Note that a small deviation is expected from the discrepancy between a MWPM decoder and a maximum-likelihood decoder [15]. With the parameters used in this work, we do not observe any loss of fidelity when we stop accounting for the difference in error rates between ancilla states. We account this to the large error contribution from photon noise and gate infidelity on the ancilla qubits, which do not have this asymmetry. We further note that we operate in a regime of large ancilla error; as described in the text this makes the system counter-intuitively less sensitive to ancilla noise. In systems where this is not the case, it could be that accounting for ancilla asymmetry provides a useful computational method to improve ϵ_L.

Appendix D: Implementation of a look-up table decoder

In [13], the authors describe a decoding scheme specific to Surface-17, which is optimized to be implementable with limited computational resources in a short cycle time. This decoding scheme works by using a short decision tree to connect errors to each other in a style similar to blossom. Indeed, this scheme is equivalent to a blossom decoder with all horizontal, vertical and diagonal weights equal [13]. As such, we have implemented the new weights in the blossom decoder rather than utilizing the exact method given.

Appendix E: Details of lowest-order approximation

We detail the approximation made to study Surface-49 in Sec. II C. Note that this calculation is only for X errors, which are measured by the Z ancillas. This implies that our approximation should attempt to realize the result of blossom, rather than the decoder upper bound.

We begin with the $\mathcal{G}_A$ graph defined in App. C. In the absence of correlated errors that cause more than two error signals, any experiment can be approximately described by choosing a set $S_e \subset E_A$ of edges on the graph and assuming the errors that correspond to these edges have occurred. Each ancilla measurement corresponds to a vertex in $\mathcal{G}_A$, which records an error if an odd number of edges in S_e point to the vertex. Each combination M_a of ancilla measurements can be generated by multiple error sets S_e.

Formally, let us write $\mathcal{M}$ for the set of all combinations of ancilla measurements and $\mathcal{S}$ for the set of all combinations of errors (so $\mathcal{S} = 2^{E_A}$). We then define a function $\phi : \mathcal{S} \to \mathcal{M}$ that takes a combination of errors to the resultant measurement outcomes. Let us fix a logical Z operator Z_L on the surface-code fabric. Then to each $S_e \in \mathcal{S}$ we can assign a parity $p(S_e) = \pm 1$ depending on whether the product of all errors in S_e commute with Z or not. A decoding then consists of a choice of parity $p_d(M_a)$ for each $M_a \in \mathcal{M}$. Such a decoding correctly decodes $S_e \in \mathcal{S}$ if $p_d(\phi(S_e)) = p(S_e)$, and creates a logical error otherwise. The source of logical errors in a perfect decoder is then precisely the fact that we can have two error combinations $S_1, S_2 \in \mathcal{S}$ such that $\phi(S_1) = \phi(S_2)$ but $p(S_1) \neq p(S_2)$.

The above suggests a method by which a perfect decoder can be constructed. Let $\phi^{-1}(M_a) \subset \mathcal{S}$ be the set of error combinations S_e that return a measurement $M_a \in \mathcal{M}$. Then, for each error combination S_e, we can calculate the probability of this occurring:

$$r(S_e) = \prod_{e \in S_e} p_e \prod_{e \notin S_e} (1 - p_e). \quad (E1)$$

The optimal choice of $p_d(M_a)$ is the one maximizing

$$\sum_{S_e \in \phi^{-1}(M_a), p(S_e) = p_d(M_a)} r(S_e), \quad (E2)$$

and the error rate of such a decoder (for the entire experiment) can be calculated as

$$\sum_{M_a \in \mathcal{M}} \min \left(\sum_{S_e \in \phi^{-1}(M_a)} \delta_{p(S_e),+1} \; r(S_e) \, , \right.$$
$$\left. \sum_{S_e \in \phi^{-1}(M_a)} \delta_{p(S_e),-1} \; r(S_e) \right). \quad (E3)$$

Unfortunately, the above function cannot be evaluated exactly; the number of error combinations S_e is approximately 2^{200} for 4 cycles of Surface-49. Our goal instead is to approximate this to the lowest order in the physical qubit error rate.

Within the bulk of a long experiment, the graph $\mathcal{G}_A$ is translationally invariant in time. This implies that we can assign to each S_e a time step $t(S_e)$, being the earliest time of the first error measurement observed (in $\phi(S_e)$). Then, it suffices to calculate Eq. (E3) for one time step to immediately obtain the logical error rate. We also only consider the result of Z-type ancilla measurements in M_a, which equates to ignoring the correlations caused by Y-errors. Next, we make a simplifying approximation by only considering error combinations S_e that cause a logical error. That is, we assume that if a logical error occurs, it occurs regardless of future physical errors in the system. This allows us to replace the probability function $r(S_e)$ with

$$\bar{r}(S_e) = \prod_{e \in S_e} p_e. \quad (E4)$$

To calculate results for Fig. 5, we evaluate Eq. (E3) for S_e containing up to $(d+1)/2$ errors (for distance d). For an S_e containing $(d+1)/2$ physical errors to cause a logical error, it must spread directly across the chain, and it cannot spread further than $(d+3)/2$ in time. As such, we only take S_e that fulfill these criteria. For each such S_e, we calculate $r(S_e)$, $\phi(S_e)$ and $p(S_e)$, and bin groups of S_e by $\phi(S_e)$. The minimization calculation is then performed by summing over the calculated bins, as opposed to determining $p_d(S_e)$ by a given decoder.

[1] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jef-

frey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Phys. Rev. X **6**, 031007 (2016).

[2] R. Barends, L. Lamata, J. Kelly, L. García-Álvarez, A. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. OMalley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, E. Solano, and J. Martinis, Nat. Commun. **6**, 7654 (2015).

[3] N. K. Langford, R. Sagastizabal, M. Kounalakis, C. Dickel, A. Bruno, F. Luthi, D. J. Thoen, A. Endo, and L. DiCarlo, arXiv:1610.10065 (2016).

[4] J. Kelly, R. Barends, A. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, *et al.*, Nature **519**, 66 (2015).

[5] D. Ristè, S. Poletto, M. Z. Huang, A. Bruno, V. Vesterinen, O. P. Saira, and L. DiCarlo, Nat. Commun. **6**, 6983 (2015).

[6] A. D. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow, Nat. Commun. **6**, 6979 (2015).

[7] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Nature **536**, 441 (2016).

[8] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, J. M. Martinis, and H. Neven, arXiv:1608.00263 (2016).

[9] J. M. Martinis, npj Quantum Inf. **1**, 15005 (2015).

[10] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Phys. Rev. A **86**, 032324 (2012).

[11] A. J. Landahl, J. T. Anderson, and P. R. Rice, arXiv:1108.5738 (2011).

[12] T. J. Yoder and I. H. Kim, arXiv:1612.04795 (2016).

[13] Y. Tomita and K. M. Svore, Phys. Rev. A **90**, 062320 (2014).

[14] A. G. Fowler, A. M. Stephens, and P. Groszkowski, Phys. Rev. A **80**, 052312 (2009).

[15] B. Heim, K. M. Svore, and M. B. Hastings, arXiv:1609.06373 (2016).

[16] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **69**, 062320 (2004).

[17] C. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, New J. Phys. **14**, 123011 (2012).

[18] R. Versluis, S. Poletto, N. Khammassi, N. Haider, D. J. Michalak, A. Bruno, K. Bertels, and L. DiCarlo, arXiv:1612.08208 (2016).

[19] C. C. Bultink, M. A. Rol, T. E. O'Brien, X. Fu, B. C. S. Dikken, C. Dickel, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, R. N. Schouten, and L. DiCarlo, Phys. Rev. Appl. **6**, 034008 (2016).

[20] M. A. Rol, C. C. Bultink, T. E. O'Brien, S. R. de Jong, L. S. Theis, X. Fu, F. Luthi, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, D. Deurloo, R. N. Schouten, F. K. Wilhelm, and L. DiCarlo, arXiv:1611.04815 (2016).

[21] S. Asaad, C. Dickel, S. Poletto, A. Bruno, N. K. Langford, M. A. Rol, D. Deurloo, and L. DiCarlo, npj Quantum Inf. **2**, 16029 (2016).

[22] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potocnik, Y. Salathe, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff,

[23] Please visit https://github.com/brianzi/quantumsim.

[24] A distance-d code with majority voting alone should exhibit a $(d+1)/2$-order decay.

[25] A. Frisk Kockum, L. Tornberg, and G. Johansson, Phys. Rev. A **85**, 052318 (2012).

[26] D. Ristè and L. DiCarlo, in *Superconducting devices in quantum optics* (Springer, 2015).

[27] H. Paik, A. Mezzacapo, M. Sandberg, D. T. McClure, B. Abdo, A. D. Córcoles, O. Dial, D. F. Bogorin, B. L. T. Plourde, M. Steffen, A. W. Cross, J. M. Gambetta, and J. M. Chow, Phys. Rev. Lett. **117**, 250502 (2016).

[28] S. Gustavsson, F. Yan, G. Catelani, J. Bylander, A. Kamal, J. Birenbaum, D. Hover, D. Rosenberg, G. Samach, A. P. Sears, S. J. Weber, J. L. Yoder, J. Clarke, A. J. Kerman, F. Yoshihara, Y. Nakamura, T. P. Orlando, and W. D. Oliver, Science **354**, 1573 (2016), http://science.sciencemag.org/content/354/6319/1573.full.pdf.

[29] B. M. Terhal, Rev. Mod. Phys. **87**, 307 (2015).

[30] A. G. Fowler, D. Sank, J. Kelly, R. Barends, and J. M. Martinis, arXiv:1405.1454 (2014).

[31] N. Delfosse and J. P. Tillich, in *2014 IEEE International Symposium on Information Theory* (2014) pp. 1071–1075.

[32] A. G. Fowler, arXiv:1310.0863 (2013).

[33] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, arXiv:1310.4492 (2013).

[34] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Nature **508**, 500 (2014).

[35] Z. Chen, Y. Wang, T. Li, L. Tian, Y. Qiu, K. Inomata, F. Yoshihara, S. Han, F. Nori, J. Tsai, *et al.*, arXiv preprint arXiv:1602.01584 (2016).

[36] A. G. Fowler, Phys. Rev. A **88**, 042308 (2013).

[37] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis, Caltech (1997).

[38] We thank Barbara Terhal for providing this derivation.

[39] A. P. Sears, A. Petrenko, G. Catelani, L. Sun, H. Paik, G. Kirchmair, L. Frunzio, L. I. Glazman, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. B **86**, 180504 (2012).

[40] D. Ristè, C. C. Bultink, M. J. Tiggelman, R. N. Schouten, K. W. Lehnert, and L. DiCarlo, Nat. Commun. **4**, 1913 (2013).

[41] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, arXiv:1605.07674 (2016).

[42] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature **460**, 240 (2009).

[43] F. Yan, S. Gustavsson, A. Kamal, J. Birenbaum, A. P. Sears, D. Hover, G. Samach, T. J. Gudmundsen, J. L. Yoder, T. P. Orlando, J. Clarke, A. J. Kerman, and W. D. Oliver, arXiv:1508.06299 (2015).

[44] C. M. Quintana, Y. Chen, D. Sank, A. G. Petukhov, T. C. White, D. Kafri, B. Chiaro, A. Megrant, R. Barends, B. Campbell, Z. Chen, A. Dunsworth, A. G. Fowler, R. Graff, E. Jeffrey, J. Kelly, E. Lucero, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, P. Roushan, A. Shabani, V. N. Smelyanskiy, A. Vainsencher, J. Wenner, H. Neven, and J. M. Martinis, Phys. Rev. Lett. **118**,

057702 (2017).

[45] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 060501 (2012).

[46] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (2000).

[47] D. Ristè, J. G. van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, Phys. Rev. Lett. **109**, 050507 (2012).

[48] M. A. Castellanos-Beltran, K. D. Irwin, G. C. Hilton, L. R. Vale, and K. W. Lehnert, Nat. Phys. **4**, 929 (2008).

[49] J. Edmonds, Canad. J. Math. **17**, 449 (1965).

[50] A. G. Fowler, arXiv:quant-ph/1210.4626 (2012).