

# Logical-qubit operations in an error-detecting surface code

## Jorge Marques

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Boris Varbanov

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Miguel Moreira

Delft University of Technology, <https://orcid.org/0000-0002-7100-8013>

## Hany Ali

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Nandini Muthusubramanian

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands  
<https://orcid.org/0000-0003-0332-0280>

## Christos Zachariadis

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Francesco Battistel

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands  
<https://orcid.org/0000-0003-4800-2518>

## Marc Beekman

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Nadia Haider

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Wouter Vlothuizen

QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

## Alessandro Bruno

Qutech Advanced Research Center, Delft University of Technology

## Barbara Terhal

Delft University

## Leonardo DiCarlo (✉ [L.DiCarlo@tudelft.nl](mailto:L.DiCarlo@tudelft.nl))

Delft University of Technology

---

## Article

**Keywords:** logical qubits, logical operations, error detection cycles, surface code

**Posted Date:** April 8th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-380732/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Physics on December 16th, 2021.

See the published version at <https://doi.org/10.1038/s41567-021-01423-9>.

# Logical-qubit operations in an error-detecting surface code

J. F. Marques,<sup>1,2</sup> B. M. Varbanov,<sup>1</sup> M. S. Moreira,<sup>1,2</sup> H. Ali,<sup>1,2</sup> N. Muthusubramanian,<sup>1,2</sup> C. Zachariadis,<sup>1,2</sup> F. Battistel,<sup>1</sup> M. Beekman,<sup>1,3</sup> N. Haider,<sup>1,3</sup> W. Vlothuisen,<sup>1,3</sup> A. Bruno,<sup>1,2</sup> B. M. Terhal,<sup>1,4</sup> and L. DiCarlo<sup>1,2,\*</sup>

<sup>1</sup>*QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands*

<sup>2</sup>*Kavli Institute of Nanoscience, Delft University of Technology,  
P.O. Box 5046, 2600 GA Delft, The Netherlands*

<sup>3</sup>*Netherlands Organisation for Applied Scientific Research (TNO),  
P.O. Box 96864, 2509 JG The Hague, The Netherlands*

<sup>4</sup>*JARA Institute for Quantum Information, Forschungszentrum Juelich, D-52425 Juelich, Germany*

(Dated: March 20, 2021)

Future fault-tolerant quantum computers will require storing and processing quantum data in logical qubits. We realize a suite of logical operations on a distance-two logical qubit stabilized using repeated error detection cycles. Logical operations include initialization into arbitrary states, measurement in the cardinal bases of the Bloch sphere, and a universal set of single-qubit gates. For each type of operation, we observe higher performance for fault-tolerant variants over non-fault-tolerant variants, and quantify the difference through detailed characterization. In particular, we demonstrate process tomography of logical gates, using the notion of a logical Pauli transfer matrix. This integration of high-fidelity logical operations with a scalable scheme for repeated stabilization is a milestone on the road to quantum error correction with higher-distance superconducting surface codes.

## INTRODUCTION

Two key capabilities will distinguish an error-corrected quantum computer from present-day noisy intermediate-scale quantum (NISQ) processors [1]. First, it will initialize, transform, and measure quantum information encoded in logical qubits rather than physical qubits. A logical qubit is a highly entangled two-dimensional subspace in the larger Hilbert space of many more physical qubits. Second, it will use repetitive quantum parity checks to discretize, signal, and (with aid of a decoder) correct errors occurring in the constituent physical qubits without destroying the encoded information [2]. Provided the incidence of physical errors is below a code-specific threshold and the quantum circuits for logical operations and stabilization are fault-tolerant, the logical error rate can be exponentially suppressed by increasing the distance (redundancy) of the quantum error correction (QEC) code employed [3]. At present, the exponential suppression for specific physical qubit errors (bit-flip or phase-flip) has been experimentally demonstrated [4, 5] for repetition codes [6–8].

Leading experimental quantum platforms have taken key steps towards implementing QEC codes protecting logical qubits from general physical qubit errors. In particular, trapped-ion systems have demonstrated logical-level initialization, gates and measurements for single logical qubits in the Calderbank-Shor-Steane [9] and Bacon-Shor [10] codes. Most recently, entangling operations between two logical qubits have been demonstrated in the surface code using lattice surgery [11].

However, except for smaller-scale experiments using two ion species [12], trapped-ion experiments in QEC have so far been limited to a single round of stabilization.

In parallel, taking advantage of highly-non-demolition measurement in circuit quantum electrodynamics [13], superconducting circuits have taken key strides in repetitive stabilization of two-qubit entanglement [14, 15] and logical qubits. Quantum memories based on 3D-cavity logical qubits in cat [16, 17] and Gottesman-Kitaev-Preskill [18] codes have crossed the memory break-even point. Meanwhile, monolithic architectures have focused on logical qubit stabilization in a surface code realized with a 2D lattice of transmon qubits. Currently, the surface code [19] is the most attractive QEC code for solid-state implementation owing to its practical nearest-neighbor connectivity requirement and high error threshold. Recent experiments [5, 20] have demonstrated repetitive stabilization by post-selection in a surface code which, owing to its small size, is capable of quantum error detection but not correction.

We demonstrate a complete suite of logical-qubit operations for this small (distance-2) surface code while preserving multi-round stabilization. Our logical operations span initialization anywhere on the logical Bloch sphere, measurement in all cardinal bases, and a universal set of single-logical-qubit gates. For each type of operation, we quantify the increased performance of fault-tolerant variants over non-fault-tolerant ones. We introduce the notion of a logical Pauli transfer matrix to describe a logical gate, analogous to the procedure commonly used to describe gates on physical qubits [21].

Finally, we compare the performance of two scalable, fault-tolerant stabilizer measurement schemes compatible with our quantum hardware architecture [22].

The distance-2 surface code (Fig. 1a) uses four data qubits ( $D_1$  through  $D_4$ ) to encode one logical qubit, whose two-dimensional codespace is the even-parity (i.e., eigenvalue +1) subspace of the stabilizer set

$$\mathcal{S} = \{Z_{D_1}Z_{D_3}, X_{D_1}X_{D_2}X_{D_3}X_{D_4}, Z_{D_2}Z_{D_4}\}. \quad (1)$$

This codespace has logical Pauli operators

$$Z_L = Z_{D_1}Z_{D_2}, Z_{D_3}Z_{D_4}, Z_{D_1}Z_{D_4}, \text{ and } Z_{D_2}Z_{D_3}, \quad (2)$$

$$X_L = X_{D_1}X_{D_3} \text{ and } X_{D_2}X_{D_4}, \quad (3)$$

that anti-commute with each other and commute with  $\mathcal{S}$ , and logical computational basis

$$|0_L\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle), \quad (4)$$

$$|1_L\rangle = \frac{1}{\sqrt{2}}(|0101\rangle + |1010\rangle). \quad (5)$$

Measuring the stabilizers using three ancilla qubits ( $A_1$ ,  $A_2$  and  $A_3$  in Fig. 1a) allows detection of all individual physical-qubit errors. Such errors change the outcome of one or more stabilizers to  $m = -1$ . However, no error syndrome combination is unique to a single error. For instance, a phase flip in any one data qubit triggers the same syndrome:  $m_{A_2} = -1$ . Consequently, this code cannot be used to correct such errors. We thus perform state stabilization by post-selecting runs in which no error is detected by the stabilizer measurements in any cycle. In this error-detection context, an operation is fault-tolerant if any single-fault produces a non-trivial syndrome and can therefore be post-selected out [23].

## RESULTS

### Stabilizer measurements

Achieving high performance in a code hinges on performing projective quantum parity (stabilizer) measurements with high assignment fidelity and low additional backaction. We implement each of the stabilizers in  $\mathcal{S}$  using a standard indirect-measurement scheme [24, 25] with a dedicated ancilla. As a fidelity metric, we measure the average probability to correctly assign the parity  $Z_{D_1}Z_{D_3}$ ,  $Z_{D_1}Z_{D_2}Z_{D_3}Z_{D_4}$  and  $Z_{D_1}Z_{D_4}$  of physical computational states of the data-qubit register, finding 94.2%, 86.1% and 97.2%, respectively (see Fig. S2).

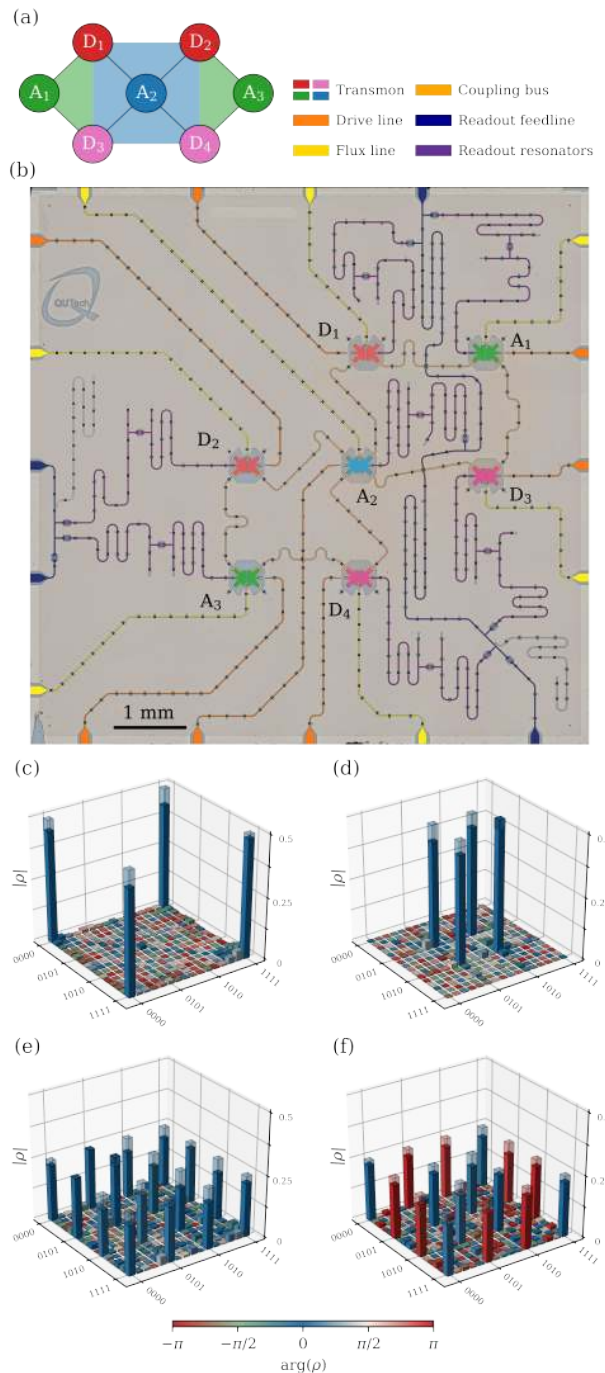


Figure 1. **Surface-7 quantum processor and initialization of logical cardinal states.** (a) Distance-two surface code. (b) Optical image of the quantum hardware with added false-color to emphasize different circuit elements. (c-f) Estimated physical density matrices,  $\rho$ , after targeting the preparation of the logical cardinal states  $|0_L\rangle$  (c),  $|1_L\rangle$  (d),  $|+_L\rangle$  (e) and  $|-_L\rangle$  (f). Each state is measured after preparing the data qubits in  $|0000\rangle$ ,  $|1010\rangle$ ,  $|++++\rangle$  and  $|++--\rangle$ , respectively. The ideal target state density matrix is shown in the shaded wireframe.

### Logical state initialization using stabilizer measurements

A practical means to quantify the backaction of stabilizer measurements is using them to initialize logical states. As proposed in Ref. 20, we can prepare arbitrary logical states by first initializing the data-qubit register in the product state

$$|\psi\rangle = (C_{\theta/2}|0\rangle + S_{\theta/2}|1\rangle)|0\rangle(C_{\theta/2}|0\rangle + S_{\theta/2}e^{i\phi}|1\rangle)|0\rangle \quad (6)$$

using single-qubit rotations  $R_y^\theta$  on  $D_1$  and  $R_\phi^\theta$  on  $D_3$  acting on  $|0000\rangle$  ( $C_\alpha = \cos \alpha$  and  $S_\alpha = \sin \alpha$ ). A follow-up round of stabilizer measurements ideally projects the four-qubit state onto the logical state

$$|\psi_L\rangle = \left( C_{\theta/2}^2|0_L\rangle + S_{\theta/2}^2e^{i\phi}|1_L\rangle \right) / \sqrt{C_{\theta/2}^4 + S_{\theta/2}^4} \quad (7)$$

with probability

$$P = \frac{1}{2} \left( C_{\theta/2}^4 + S_{\theta/2}^4 \right). \quad (8)$$

We use this procedure to target initialization of the logical cardinal states  $|0_L\rangle$ ,  $|1_L\rangle$ ,  $|+_L\rangle = (|0_L\rangle + |1_L\rangle)/\sqrt{2}$ , and  $|-_L\rangle = (|0_L\rangle - |1_L\rangle)/\sqrt{2}$ . For the first two states, the procedure is fault-tolerant according to the definition above. We characterize the produced states using full four-qubit state tomography including readout calibration and maximum-likelihood estimation (MLE) (Fig. 1c-f). The fidelity  $F_{4Q}$  to the ideal four-qubit target states is  $90.0 \pm 0.3\%$ ,  $92.9 \pm 0.2\%$ ,  $77.3 \pm 0.5\%$ , and  $77.1 \pm 0.5\%$ , respectively. For each state, we can extract a logical fidelity  $F_L$  by further projecting the obtained four-qubit density matrix onto the codespace [20], finding  $99.83 \pm 0.08\%$ ,  $99.97 \pm 0.04\%$ ,  $96.82 \pm 0.55\%$ , and  $95.54 \pm 0.55\%$ , respectively (see Methods). This sharp increase from  $F_{4Q}$  to  $F_L$  demonstrates that the vast majority of errors introduced by the parity check are weight-1 and detectable. A simple modification makes the initialization of  $|+_L\rangle$  ( $|-_L\rangle$ ) also fault-tolerant: initialize the data-qubit register in a different product state, namely  $|++++\rangle$  ( $|++--\rangle$ ), before performing the stabilizer measurements. With this modification,  $F_{4Q}$  increases to  $85.4 \pm 0.3\%$  ( $84.6 \pm 0.3\%$ ) and  $F_L$  to  $99.78 \pm 0.09\%$  ( $99.64 \pm 0.17\%$ ), matching the performance achieved when targetting  $|0_L\rangle$  and  $|1_L\rangle$ .

### Logical measurement of arbitrary states

A key feature of a code is the ability to measure logical operators. In the surface code, we can measure  $X_L$  ( $Z_L$ ) fault-tolerantly, albeit destructively, by

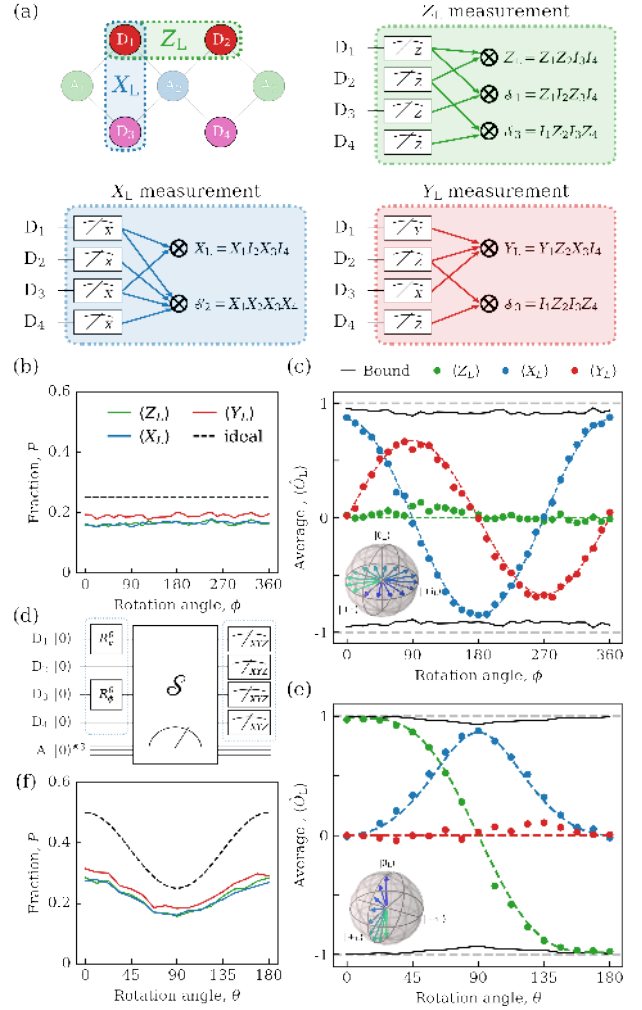


Figure 2. **Arbitrary logical-state initialization and measurement in the logical cardinal bases.** (a) Assembly of data-qubit measurements used to evaluate logical operators  $Z_L$ ,  $X_L$  and  $Y_L$  with additional error detection. (d) Initialization of logical states using the procedure described in Eq. 6. (c, e)  $Z_L$ ,  $X_L$  and  $Y_L$  logical measurement results as a function of the gate angles  $\phi$  (c) and  $\theta$  (e). The colored dashed curves show a fit of the analytical prediction based on Eqs. 9 and 11 to the data and the dark curve denotes a bound based on the measured  $F_L$  of each state. (b, f) Total fraction  $P$  of post-selected data as a function of the input angle for each logical measurement. The dashed curve shows the ideal fraction given by Eq. 8.

simultaneously measuring all data qubits in the  $X$  ( $Z$ ) basis to obtain a string of data-qubit outcomes (each +1 or -1). The value assigned to the logical operator is the computed product of data-qubit outcomes as prescribed by Eq. 3 (2). Additionally, the outcome string is used to compute a value for the stabilizer(s)

$X_{D1}X_{D2}X_{D3}X_{D4}$  ( $Z_{D1}Z_{D3}$  and  $Z_{D2}Z_{D4}$ ), enabling a final step of error detection (Fig. 2a). Measurement of  $Y_L = +iX_LZ_L = Y_{D1}Z_{D2}X_{D3}$  is not fault-tolerant. However, we lower the logical assignment error by also measuring  $D_4$  in the  $Z$  basis to compute a value for  $Z_{D2}Z_{D4}$  and thereby detect bit-flip errors in  $D_2$  and  $D_4$ .

We demonstrate  $Z_L$ ,  $X_L$  and  $Y_L$  measurements on logical states prepared on two orthogonal planes of the logical Bloch sphere. Setting  $\theta = \pi/2$  and sweeping  $\phi$ , we ideally prepare logical states on the equator (Fig. 2d)

$$|\psi_L\rangle = (|0_L\rangle + e^{i\phi}|1_L\rangle) / \sqrt{2}. \quad (9)$$

We measure the produced states in the  $Z_L$ ,  $X_L$  and  $Y_L$  bases and obtain experimental averages  $\langle Z_L \rangle$ ,  $\langle X_L \rangle$  and  $\langle Y_L \rangle$ . As expected, we observe sinusoidal oscillations in  $\langle X_L \rangle$  and  $\langle Y_L \rangle$  and near-zero  $\langle Z_L \rangle$ . We extract logical assignment fidelities  $F_L^R$  for  $X_L$  and  $Y_L$  from  $\langle X_L \rangle$  and  $\langle Y_L \rangle$  at the cardinal states accounting for initialization error given by  $F_L$ :

$$(2F_L^R - 1)(2F_L - 1) = \max|\langle O_L \rangle|, \quad O \in \{X, Y\}. \quad (10)$$

We find  $F_L^R = 95.8\%$  for  $X_L$  and  $87.5\%$  for  $Y_L$ , which manifests the non-fault-tolerant nature of  $Y_L$  measurement. A second manifestation is the higher fraction  $P$  of post-selected data in this case (Fig. 2b).

Setting  $\phi = 0$  and sweeping  $\theta$ , we then prepare logical states on the  $X_L$ - $Z_L$  plane of the logical Bloch sphere (Fig. 2e), ideally

$$|\psi_L\rangle = (C_{\theta/2}^2|0_L\rangle + S_{\theta/2}^2|1_L\rangle) / \sqrt{C_{\theta/2}^4 + S_{\theta/2}^4}. \quad (11)$$

Note that due to the changing overlap of the initial product state with the codespace,  $P$  is now a function of  $\theta$  (Eq. 8). Using the same procedure as above, we extract  $F_L^R = 99.4\%$  for  $Z_L$  and  $96.4\%$  for  $X_L$ . Although both measurements are fault-tolerant,  $F_L^R$  is higher for  $Z_L$ . This arises because the  $Z_L$  measurement is only vulnerable to vertical double bit-flip errors while the  $X_L$  measurement is vulnerable to horizontal and diagonal double phase-flip errors.

### Logical gates

Finally, we demonstrate a suite of gates enabling universal logical-qubit control (Fig. 3). Full control of the logical qubit requires a gate set comprising Clifford and non-Clifford logical gates. Some Clifford gates, like  $Z_L^\pi$  and  $X_L^\pi$ , can be implemented transversally and therefore fault-tolerantly (Fig. 3d). We perform arbitrary rotations (generally non-fault-tolerant) about the

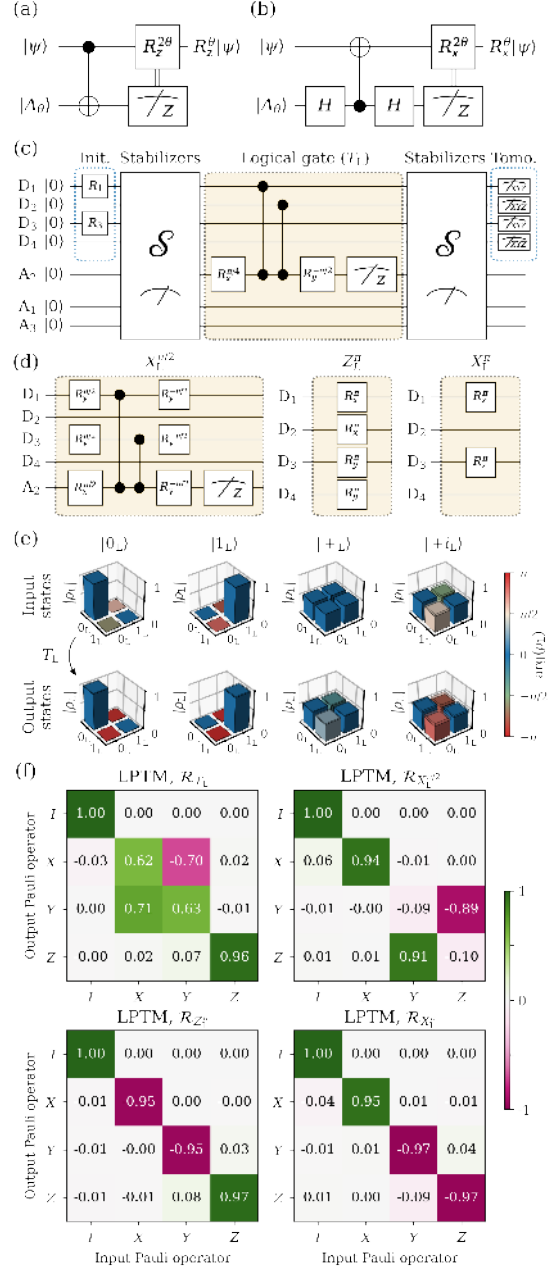


Figure 3. **Logical gates and their characterization.** (a, b) General gate-by-measurement schemes realizing arbitrary rotations around the  $Z$  (a) and  $X$  (b) axis of the Bloch sphere. (c) Process tomography experiment of the  $T_L$  gate. Input cardinal logical states are initialized using the method of Fig. 2. Output states are measured following a second round of stabilizer measurements. (d) Logical  $X_L^{\pi/2}$ ,  $Z_L^\pi$  and  $X_L^\pi$  gates compiled using our hardware-native gateset. (e) Logical state tomography of input and output states of the  $T_L$  gate. These logical density matrices are obtained by performing four-qubit tomography of the data qubits and then projecting onto the codespace. (f) Extracted logical Pauli transfer matrices.

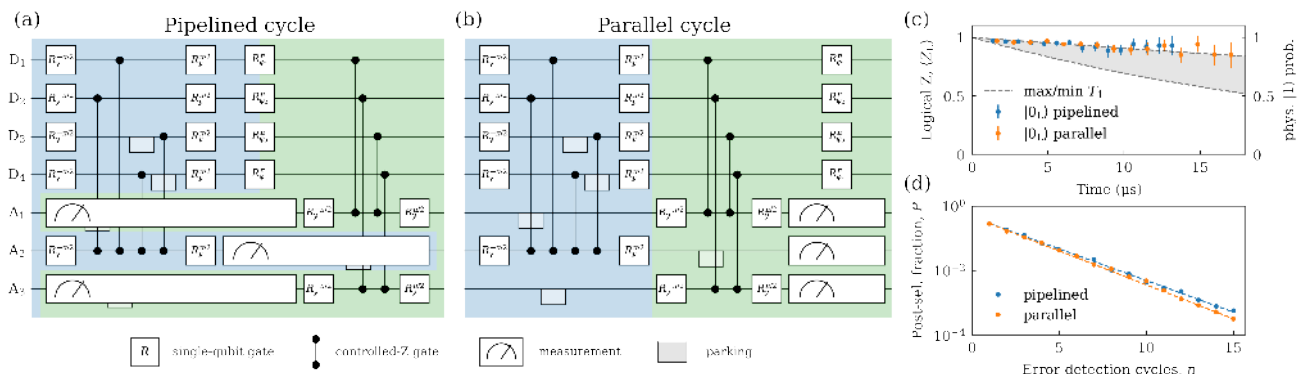


Figure 4. **Repetitive error detection using pipelined and parallel stabilizer measurement schemes.** (a, b) Gate sequences used to implement the pipelined (a) and parallel (b) stabilizer measurement schemes. Gate duration is 20 ns for single-qubit gates, 60 ns for controlled-Z (CZ) gates and parking [14, 22], and 540 ns for ancilla readout. The order of CZs in the  $X_{D1}X_{D2}X_{D3}X_{D4}$  stabilizer (blue shaded region) prevents the propagation of ancilla errors into logical qubit errors [23]. The total cycle duration for the pipelined (parallel) scheme is 840 ns (1000 ns). (c) Estimated  $Z_L$  expectation value,  $\langle Z_L \rangle$ , measured for the  $|0_L\rangle$  state versus the duration of the experiment using the pipelined (blue) and the parallel (orange) schemes. We also plot the excited-state probability (right axis) set by the maximum and minimum physical qubit  $T_1$ . (d) Post-selected fraction of data versus the number of error detection cycles  $n$  for the pipelined (blue) and parallel (orange) scheme.

$Z_L$  axis using the standard gate-by-measurement circuit [26] shown in Fig. 3a. In our case, the ancilla is physical ( $A_2$ ), while the qubit transformed is our logical qubit. The rotation angle  $\theta$  is set by the initial ancilla state  $|A_\theta\rangle = (|0\rangle + e^{i\theta}|1\rangle)/\sqrt{2}$ . Since we cannot do binary-controlled  $Z_L$  rotations, we simply post-select runs in which the measurement outcome is  $m_{A2} = +1$ . Choosing  $\theta = \pi/4$  implements the non-Clifford  $T_L = Z_L^{\pi/4}$  gate. A similar circuit (Fig. 3b) can be used to perform arbitrary rotations around the  $X_L$  axis. We compile both circuits using our hardware-native gateset (Figs. 3c,d). To assess logical-gate performance, we perform logical process tomography using the procedure illustrated in Fig. 3e for  $T_L$ . First, we initialize into each of the six logical cardinal states  $\{|0_L\rangle, |1_L\rangle, |+_L\rangle, |-_L\rangle, |+i_L\rangle, |-i_L\rangle\}$ . We characterize each actual input state by four-qubit state tomography and project to the codespace to obtain a logical density matrix. Next, we similarly characterize each output state produced by the logical gate and a second round of stabilizer measurements to detect errors occurred in the gate. Using this over-complete set of input-output logical-state pairs, combined with MLE (see Methods), we extract a logical Pauli transfer matrix (LPTM). The resulting LPTMs for the non-fault-tolerant  $T_L$  and  $X_L^{\pi/2}$  gates as well as the fault-tolerant  $Z_L^\pi$  and  $X_L^\pi$  are shown in Fig. 3e. From the LPTMs, we extract average logical gate fidelities  $F_L^G$  (Eq. 19) 97.3%, 95.6%, 97.9%, and 98.1%, respectively.

### Pipelined versus parallel stabilizer measurements

A scalable control scheme is fundamental to realize surface codes with large code distance. To this end, we now compare the performance of two schemes suitable for the quantum hardware architecture proposed in Ref. 22. These schemes are scalable in the sense that their cycle duration remains independent of code distance. The pipelined scheme interleaves the coherent operations and ancilla readout steps associated with stabilizer measurements of type  $X$  and  $Z$  by performing the coherent operations of  $X$  ( $Z$ ) type stabilizers during the readout of  $Z$  ( $X$ ) type stabilizers (Fig. 4a). The parallel scheme performs all ancilla readouts simultaneously (Fig. 4b). To compare their performance, we initialize and stabilize  $|0_L\rangle$  for up to  $n = 15$  cycles. We separately calibrate the equatorial rotation axis of refocusing pulses ( $R_{\varphi_i}^\pi$ ) in each scheme to extract the best performance in both schemes. At each  $n$ , we take data back-to-back for the two schemes in order to minimize the effect of parameter drift, repeating each experiment up to  $256 \times 10^3$  times. Figure 4c shows the  $Z_L$  measurement outcome averaged over the post-selected runs. We extract the error-detection rate  $\gamma$  from the  $n$ -dependence of the fraction of post-selected data  $P$  (Fig. 4d) using the procedure described in Methods. We observe that the error rate is slightly lower for the pipelined scheme ( $\gamma_{\text{pip}} \sim 45\%$ ), most likely due to the shorter duration of the cycle. This superiority is consistent across different input logical states (see Fig. S3)

with an average ratio  $\gamma_{\text{pip}}/\gamma_{\text{par}} \sim 97\%$ .

## DISCUSSION

We have demonstrated a suite of logical-level initialization, gate and measurement operations in a distance-2 superconducting surface code undergoing repetitive stabilizer measurements. For each type of logical operation, we have quantified the increased performance of fault-tolerant variants over non-fault-tolerant variants. Table I summarizes all the results. We can initialize the logical qubit to any point on the logical Bloch sphere, with logical fidelity surpassing Ref. 20. In addition to characterizing initialized states using full four-qubit tomography, we also demonstrate logical measurements in all logical cardinal bases. Finally, we demonstrate a universal single-qubit set of logical gates by performing logical process tomography, introducing the concept of a logical-level Pauli transfer matrix.

With a view towards implementing higher-distance surface codes using our quantum-hardware architecture [22], we have compared the performance of two scalable stabilization schemes: the pipelined and parallel measurement schemes. In this comparison, two main factors compete. On one hand, the shorter cycle time favors pipelining. On the other, the pipelining introduces extra dephasing on ancilla qubits of one type during readout of the other. The performance of both schemes is comparable, but slightly higher for the pipelined scheme. From density-matrix simulations discussed in detail in the Supplementary Material, we further understand that conventional qubit errors such as energy relaxation, dephasing and readout assignment error alone do not account for the net error-detection rate observed in the experiment (Fig. S5). We believe that the dominant error source is instead leakage to higher transmon states incurred during CZ gates. Our data (Fig. S4) shows that the error detection scheme successfully post-selects leakage errors in both the ancilla and data qubits. Learning to identify these non-qubit errors and to correct them without post-selection is the subject of ongoing research [27–29] and an outstanding challenge in the quest for quantum fault-tolerance with higher-distance superconducting surface codes.

## METHODS

### Device

We use a seven-transmon superconducting processor (Fig. 1b) featuring the quantum-hardware architecture

Logical operation	Characteristic	Logical fidelity metric	value (%)
Init.	$ 0_L\rangle$	FT	99.83
	$ 1_L\rangle$	FT	99.97
	$ +_L\rangle$	Non-FT/FT	96.82/99.78
	$ -_L\rangle$	Non-FT/FT	95.54/99.64
Meas.	$Z_L$	FT	99.4
	$X_L$	FT	96.0*
	$Y_L$	Non-FT	87.5
Gate	$Z_L^\pi$	FT	98.1
	$X_L^\pi$	FT	97.9
	$X_L^{\pi/2}$	Non-FT	95.6
	$T_L$	Non-FT	97.3

Table I. **Summary of logical initialization, measurement, and gate operations and their performance.** Fault-tolerant operations are labelled FT and non-fault tolerant ones Non-FT. \*Weighted average of values extracted from Figs. 2c,e.

proposed in Ref. 22. We employ flux-tunable transmons arranged in three frequency groups: a high-frequency group for  $D_1$  and  $D_2$ ; a middle-frequency group for  $A_1$ ,  $A_2$  and  $A_3$ ; and a low-frequency group for  $D_3$  and  $D_4$ . Each transmon is transversely coupled to its nearest neighbor using a dedicated coupling bus resonator and features an individual microwave drive line for single-qubit gates, a flux line for two-qubit gates, and a dispersively coupled readout resonator with Purcell filter for readout [15, 30]. All transmons are flux biased to their maximal frequency (i.e., flux sweetspot [31]). Qubit relaxation ( $T_1$ ) and dephasing ( $T_2$ ) times lie in the range 27–102  $\mu\text{s}$  and 55–117  $\mu\text{s}$ , respectively. Detailed information on the implementation and performance of single- and two-qubit gates can be found in Ref. 32. Device characteristics are also summarized in Table S1.

### State tomography

To perform state tomography on the prepared logical states, we measure the  $4^4 - 1$  expectation values of data-qubit Pauli observables,  $p_i = \langle \sigma_i \rangle$ ,  $\sigma_i \in \{I, X, Y, Z\}^{\otimes 4}$  (except  $I^{\otimes 4}$ ). These are used to construct the density matrix

$$\rho = \sum_{i=0}^{4^4-1} \frac{p_i \sigma_i}{2^4} \quad (12)$$

with  $p_0 = 1$ , corresponding to  $\sigma_0 = I^{\otimes 4}$ . Due to statistical uncertainty in the measurement, the constructed state,  $\rho$ , might lack the physicality characteristic of a density matrix, that is,  $\text{Tr}(\rho) = 1$  and  $\rho \geq 0$ . Specifically,  $\rho$  might not satisfy the latter constraint, while the former is automatically satisfied by  $p_0 = 1$ . To enforce these constraints, we use a



maximum-likelihood method [21] to find the physical density matrix,  $\rho_{\text{ph}}$ , that is closest to the measured state, where closeness is defined in terms of best matching the measurement results. We thus minimize the cost function  $\sum_{i=0}^{4^A-1} |p_i - \text{Tr}(\rho_{\text{ph}}\sigma_i)|^2$ , subject to  $\text{Tr}(\rho_{\text{ph}}) = 1$  and  $\rho_{\text{ph}} \geq 0$ . We find the optimal  $\rho_{\text{ph}}^{\text{opt}}$  using the convex-optimization package *cvxpy* via *cvx-fit* in Qiskit [33]. The fidelity to a target pure state,  $|\psi\rangle$ , is then computed as

$$F = \langle \psi | \rho_{\text{ph}}^{\text{opt}} | \psi \rangle. \quad (13)$$

One can further project  $\rho_{\text{ph}}$  onto the codespace to obtain a logical state  $\rho_{\text{L}}$  using

$$\rho_{\text{L}} = \frac{1}{2} \sum_i \frac{\text{Tr}(\rho_{\text{ph}}\sigma_i^{\text{L}})}{\text{Tr}(\rho_{\text{ph}}I_{\text{L}})} \sigma_i^{\text{L}}, \quad \sigma_i^{\text{L}} \in \{I_{\text{L}}, X_{\text{L}}, Y_{\text{L}}, Z_{\text{L}}\} \quad (14)$$

where  $I_{\text{L}}$  is the projector onto the codespace. Here, we can compute the logical fidelity  $F_{\text{L}}$  using Eq. 13.

### Process tomography in the codespace

A general single-qubit gate can be described [21] by a Pauli transfer matrix (PTM)  $\mathcal{R}$  that maps an input state described by  $p_i = \langle \sigma_i \rangle$ ,  $\sigma_i \in \{I, X, Y, Z\}$ , with  $p_0 = 1$ , to an output state  $p'_j$ :

$$p'_j = \sum_i \mathcal{R}_{ij} p_i. \quad (15)$$

To construct  $\mathcal{R}$  in the codespace, we use an overcomplete set of input states,  $\{|0_{\text{L}}\rangle, |1_{\text{L}}\rangle, |+\text{L}\rangle, |-\text{L}\rangle, |+i_{\text{L}}\rangle, |-i_{\text{L}}\rangle\}$ , and their corresponding output states and perform linear inversion. The input and output logical states are characterized using state tomography of the data qubits to find the four-qubit state  $\rho$ , which is then projected to the codespace using:

$$p_i^{\text{L}} = \frac{\text{Tr}(\rho\sigma_i^{\text{L}})}{\text{Tr}(\rho I_{\text{L}})}, \quad \sigma_i^{\text{L}} \in \{I_{\text{L}}, X_{\text{L}}, Y_{\text{L}}, Z_{\text{L}}\}, \quad (16)$$

We find that all the measured logical states already satisfy the constraints of a physical density matrix. This is likely to happen as one-qubit states that are not very pure usually lie within the Bloch sphere even within the uncertainty in the measurement. The constructed LPTM, however, might not satisfy the constraints of a physical quantum channel, that is, trace preservation and complete positivity (TPCP). These are better expressed by switching from the PTM representation to the Choi representation. The Choi state  $\rho^{\mathcal{R}}$  can be computed as

$$\rho^{\mathcal{R}} = \frac{1}{4} \sum_{i,j} \mathcal{R}_{ij} \sigma_j^T \otimes \sigma_i, \quad (17)$$

where the first tensor-product factor corresponds to an auxiliary subsystem. The TPCP constraints are  $\text{Tr}(\rho_{\text{ph}}^{\mathcal{R}}) = 1$ ,  $\rho_{\text{ph}}^{\mathcal{R}} \geq 0$  and  $\text{Tr}_1(\rho_{\text{ph}}^{\mathcal{R}}) = 1/2$ , where  $\text{Tr}_1$  is the partial trace over the auxiliary subsystem. In other words,  $\rho_{\text{ph}}^{\mathcal{R}}$  is a density matrix satisfying an extra constraint. We then find the optimal  $\rho_{\text{ph}}^{\mathcal{R},\text{opt}}$  using the same convex-optimization methods as for state tomography and adding this extra constraint [21, 34]. We compute the corresponding LPTM via

$$(\mathcal{R}_{\text{ph}}^{\text{opt}})_{ij} = \text{Tr}(\rho_{\text{ph}}^{\mathcal{R},\text{opt}} \sigma_j^T \otimes \sigma_i). \quad (18)$$

and the average logical gate fidelity using

$$F_{\text{L}}^{\text{G}} = \frac{\text{Tr}(\mathcal{R}_{\text{ideal}}^{\dagger} \mathcal{R}_{\text{ph}}^{\text{opt}}) + 2}{6}, \quad (19)$$

where  $\mathcal{R}_{\text{ideal}}$  is the LPTM of the ideal target gate.

### Extraction of error-detection rate

The fraction of post-selected data  $P$  in the repetitive error detection experiment (Fig. 4b) decays exponentially with the number of cycles  $n$ . This is consistent with a constant error-detection rate per cycle  $\gamma$ . We extract this rate by fitting the function

$$P(n) = A(1 - \gamma)^n. \quad (20)$$

---

\* Corresponding author: L.dicarlo@tudelft.nl

- [1] Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- [2] Terhal, B. M. Quantum error correction for quantum memories. *Rev. Mod. Phys.* **87**, 307–346 (2015)
- [3] Martinis, J. M. Qubit metrology for building a fault-tolerant quantum computer. *npj Quantum Inf.* **1**, 15005 (2015).
- [4] Kelly, J. *et al.* State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).
- [5] Chen, Z. *et al.* Exponential suppression of bit or phase flip errors with repetitive error correction. Preprint at <http://arXiv.org/abs/2102.06132> (2021).
- [6] Ristè, D. *et al.* Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nat. Commun.* **6**, 6983 (2015).
- [7] Cramer, J. *et al.* Repeated quantum error correction on a continuously encoded qubit by real-time feedback. *Nat. Commun.* **5**, 11526 (2016).
- [8] Ristè, D. *et al.* Real-time processing of stabilizer measurements in a bit-flip code. *npj Quantum Information* **6**, 71 (2020).
- [9] Nigg, D. *et al.* Quantum computations on a topologically encoded qubit. *Science* **345**, 302–305 (2014).

- [10] Egan, L. *et al.* Fault-tolerant operation of a quantum error-correction code. Preprint at <http://arXiv.org/abs/2009.11482> (2021).
- [11] Erhard, A. *et al.* Entangling logical qubits with lattice surgery. *Nature* **589**, 220–224 (2021).
- [12] Negnevitsky, V. *et al.* Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register. *Nature* **563**, 527–531 (2018).
- [13] Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
- [14] Andersen, C. K. *et al.* Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits. *npj Quantum Information* **5**, 1–7 (2019).
- [15] Bultink, C. C. *et al.* Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements. *Science Advances* **6** (2020).
- [16] Ofek, N. *et al.* Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441 (2016).
- [17] Hu, L. *et al.* Quantum error correction and universal gate set operation on a binomial bosonic logical qubit. *Nat. Phys.* (2019).
- [18] Campagne-Ibarcq, P. *et al.* Quantum error correction of a qubit encoded in grid states of an oscillator. *Nature* **584**, 368–372 (2020).
- [19] Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
- [20] Andersen, C. K. *et al.* Repeated quantum error detection in a surface code. *Nat. Phys.* **16**, 875–880.
- [21] Chow, J. M. *et al.* Universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits. *Phys. Rev. Lett.* **109**, 060501 (2012).
- [22] Versluis, R. *et al.* Scalable quantum circuit and control for a superconducting surface code. *Phys. Rev. App.* **8**, 034021 (2017).
- [23] Tomita, Y. & Svore, K. M. Low-distance surface codes under realistic quantum noise. *Phys. Rev. A* **90**, 062320 (2014).
- [24] Saira, O.-P. *et al.* Entanglement genesis by ancilla-based parity measurement in 2D circuit QED. *Phys. Rev. Lett.* **112**, 070502 (2014).
- [25] Takita, M. *et al.* Demonstration of weight-four parity measurements in the surface code architecture. *Phys. Rev. Lett.* **117**, 210505 (2016).
- [26] Aliferis, P., Gottesman, D. & Preskill, J. Quantum accuracy threshold for concatenated distance-3 codes. *Quantum Inf. Comput.* **6**, 97–165 (2005).
- [27] Varbanov, B. M. *et al.* Leakage detection for a transmon-based surface code. *npj Quantum Information* **6**, 102 (2020).
- [28] McEwen, M. *et al.* Removing leakage-induced correlated errors in superconducting quantum error correction. Preprint at <http://arXiv.org/abs/2102.06131> (2021).
- [29] Battistel, F., Varbanov, B. M. & Terhal, B. M. A hardware-efficient leakage-reduction scheme for quantum error correction with superconducting transmon qubits. Preprint at <http://arXiv.org/abs/2102.08336> (2021).
- [30] Heinsoo, J. *et al.* Rapid high-fidelity multiplexed readout of superconducting qubits. *Phys. Rev. App.* **10**, 034040 (2018).
- [31] Schreier, J. A. *et al.* Suppressing charge noise decoherence in superconducting charge qubits. *Phys. Rev. B* **77**, 180502(R) (2008).
- [32] Negirneac, V. *et al.* High-fidelity controlled-z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor. Preprint at <http://arXiv.org/abs/2008.07411> (2020).
- [33] Qiskit: An open-source framework for quantum computing (2019).
- [34] de Jong, J. *Implementation of a fault-tolerant SWAP operation on the IBM 5-qubit device.* Master’s thesis, Delft University of Technology (2019).

## ACKNOWLEDGEMENTS

We thank R. Sagastizabal, M. Sarsby and T. Stavenga for experimental assistance, and G. Calusine and W. Oliver for providing the traveling-wave parametric amplifiers used in the readout amplification chain. This research is supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the U.S. Army Research Office Grant No. W911NF-16-1-0071, and by Intel Corporation. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. B. M. V., F. B. and B. M. T. are supported by ERC Grant EQEC No. 682726.

## AUTHOR CONTRIBUTIONS

J. F. M. performed the experiment and data analysis. M. B., N. H. and L. D. C. designed the device. N. M., C. Z. and A. B. fabricated the device. J. F. M. and H. A. calibrated the device. M. S. M. and W. V. designed the control electronics. B. M. V. performed the numerical simulations and F. B. implemented the MLE method. B. M. T. supervised the theory work. J. F. M. and L. D. C. wrote the manuscript with contributions from B. M. V., F. B. and B. M. T., and feedback from all coauthors. L. D. C. supervised the project.

## COMPETING INTERESTS

The authors declare no competing interests.

**SUPPLEMENTAL MATERIAL FOR 'LOGICAL-QUBIT OPERATIONS IN AN ERROR-DETECTING SURFACE CODE'**

This supplement provides additional information in support of statements and claims made in the main text.

**DEVICE CHARACTERISTICS**

Qubit	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>
Qubit transition frequency at sweetspot, $\omega_q/2\pi$ (GHz)	6.433	6.253	4.535	4.561	5.770	5.881	5.785
Transmon anharmonicity, $\alpha/2\pi$ (MHz)	-280	—	-320	—	-290	-285	—
Readout frequency, $\omega_r/2\pi$ (GHz)	7.493	7.384	6.913	6.645	7.226	7.058	7.101
Relaxation time, $T_1$ ( $\mu$ s)	27	44	32	102	38	58	43
Ramsey dephasing time, $T_2^*$ ( $\mu$ s)	44	55	51	103	55	60	52
Echo dephasing time, $T_2$ ( $\mu$ s)	59	70	55	117	69	79	73
Best multiplexed readout fidelity, $F_{RO}$ , (%)	98.6	98.9	96.0	96.5	98.6	94.2	98.9

Table S1. **Summary of frequency, coherence and readout parameters of the seven transmons.** Coherence times are obtained using standard time-domain measurements [S1]. Note that temporal fluctuations of several  $\mu$ s are typical for these values. The multiplexed readout fidelity,  $F_{RO}$ , is the average assignment fidelity [S2] extracted from single-shot readout histograms after mitigating residual excitation using initialization by measurement and post-selection [S3, S4].

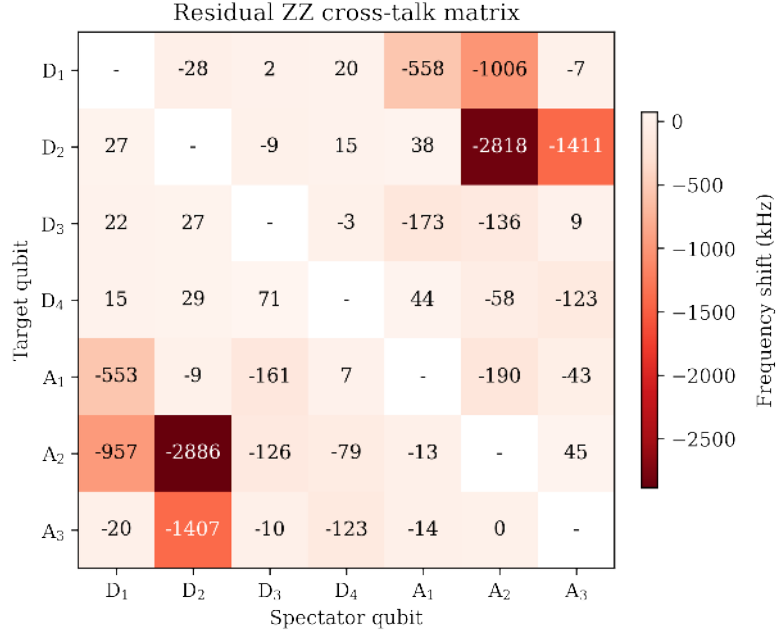


Figure S1. **Residual ZZ-coupling matrix.** Measured residual ZZ coupling between all transmon pairs at the bias point (their simultaneous flux sweetspot [S5]). Each matrix element denotes the frequency shift that the target qubit experiences due to the spectator qubit being in the excited state,  $|1\rangle$ . The procedure used for this measurement is similar to the one described in Ref. S6.

## PARITY-CHECK PERFORMANCE

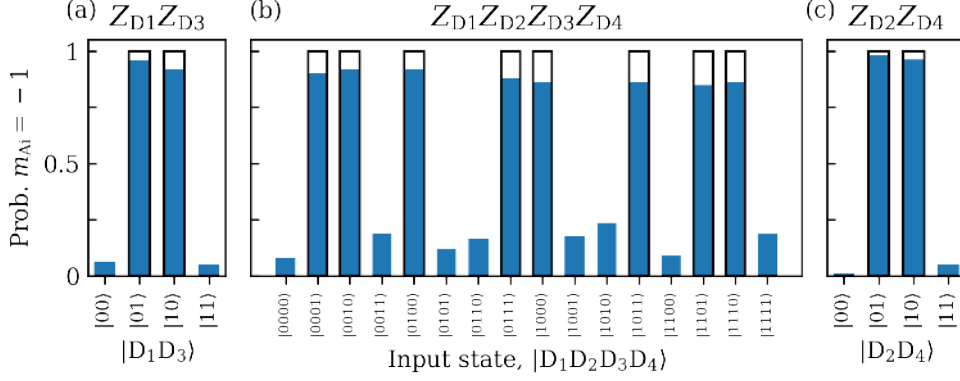


Figure S2. **Characterization of the assignment fidelity of Z-type parity checks.** (a)  $Z_{D1}Z_{D3}$ , (b)  $Z_{D1}Z_{D2}Z_{D3}Z_{D4}$ , and (c)  $Z_{D2}Z_{D4}$  parity checks implemented using  $A_1$ ,  $A_2$ , and  $A_3$ , respectively. Each parity check is benchmarked by preparing the relevant data qubits in a computational state and then measuring the probability of ancilla outcome  $m_{A_i} = -1$ . Measured (ideal) probabilities are shown as solid blue bars (black wireframe). From the measured probabilities we extract average assignment fidelities 94.2%, 86.1% and 97.2%, respectively.

## STATE STABILIZATION

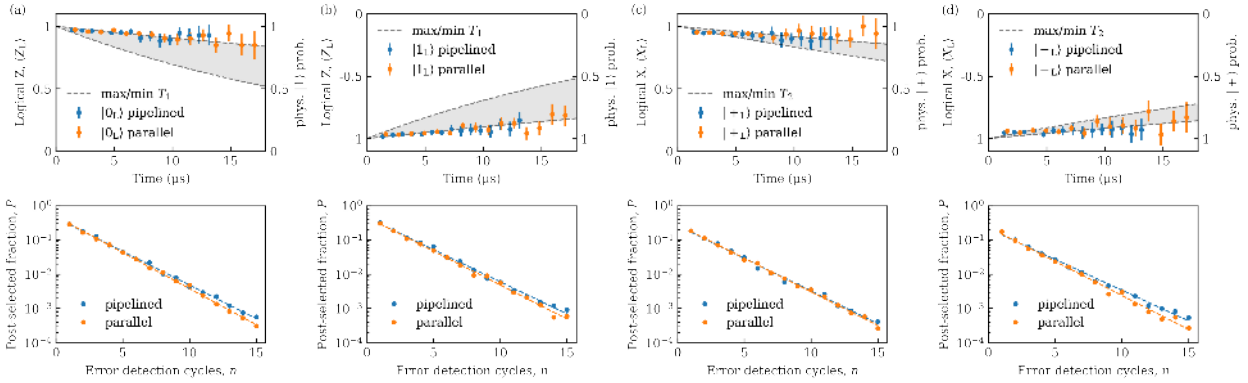


Figure S3. **Stabilization of logical cardinal states by repetitive error detection using the pipelined and parallel schemes.** From left to right, the stabilized logical states are  $|0_L\rangle$ ,  $|1_L\rangle$ ,  $|+_L\rangle$  and  $|-_L\rangle$ . For each logical state, the top panel shows the evolution of the relevant logical operator as a function of number of cycles,  $n$ , plotted versus wall-clock time. Error bars are estimated based on the statistical uncertainty given by  $P(n)$ . The shaded area indicates the range of physical qubit  $T_1$  values (a and b) and  $T_2$  values (c and d) plotted on the right-axis. Each bottom panel shows the corresponding post-selected fraction of data,  $P(n)$ .

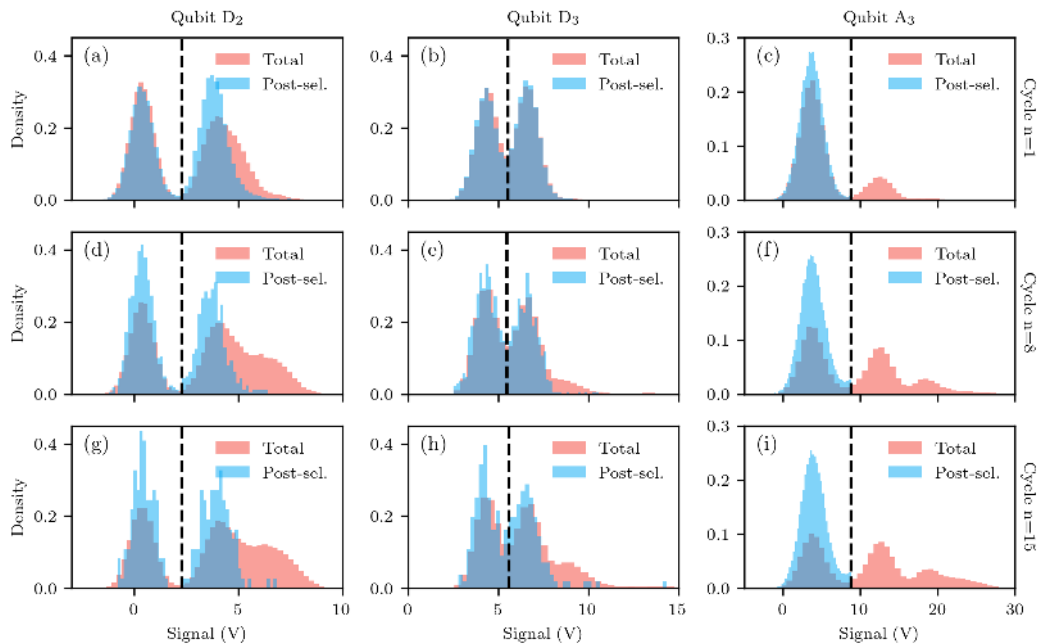


Figure S4. **Signature of transmon leakage in experimental data.** Single-shot readout histograms obtained at cycle  $n$  over all shots (red) and the post-selected shots based on detecting no error in any cycles up to  $n$  (blue) for  $D_2$  (left),  $D_3$  (middle) and  $A_3$  (right) and at cycle  $n = 1$  (top row),  $n = 8$  (middle row) and  $n = 15$  (bottom row). The dashed black lines indicate the thresholds used to discriminate  $|0\rangle$  from  $|1\rangle$ .

## NUMERICAL ANALYSIS

### Leakage in experiment

We observe a clear signature of leakage accumulation with the increasing number of error-detection cycles in the single-shot readout histograms obtained at the end of each experiment. In Fig. S4 we show examples of this accumulation for  $D_2$ ,  $D_3$  and  $A_3$  at cycles  $n = 1$ ,  $n = 8$  and  $n = 15$ . For dispersive readout, a transmon in state  $|2\rangle$  induces a different frequency shift in the readout resonator compared to state  $|0\rangle$  or  $|1\rangle$ . The increased number of data points at  $n = 8$  and  $n = 15$  shown in Fig. S4, following a Gaussian distribution with a mean and standard deviation different from those observed at  $n = 1$  is thus a clear manifestation of leakage to the higher-excited states (mostly to  $|2\rangle$ ). We believe that the dominant source of leakage in our processor are the CZ gates [S7, S8]. However, the leakage rate  $L_1$  for each gate has not been experimentally characterized, e.g., by performing leakage-modified randomized benchmarking experiments [S9, S10]. This is because our CZ tune-up procedure is performed in a parity-check block unit. This maximizes the performance of the parity-check but makes the gate unfit for randomized benchmarking protocols. We can estimate the population  $p^{\mathcal{L}}(n)$  in the leakage subspace  $\mathcal{L}$  at cycle  $n$  from the single-shot readout histograms. We perform a fit of a triple Gaussian model to the histograms from which we extract the voltage that allows for the best discrimination of  $|2\rangle$  from  $|1\rangle$  and  $|0\rangle$ . The leaked population  $p^{\mathcal{L}}(n)$  is then given by the fraction of shots declared as  $|2\rangle$  over the total number of shots. Assuming that leakage is only induced by the CZ gates (on the transmon being fluxed to perform the gate) and that each CZ gate has the same leakage rate  $L_1$ , we can use the Markovian model presented in Ref. S11 to estimate the  $L_1$  value leading to the observed population  $p^{\mathcal{L}}(n)$ . This analysis gives a  $L_1$  estimate in the approximate range 1 – 4% for most transmons. However, we do not consider these estimates to be accurate due to the low fidelity with which  $|2\rangle$  can be distinguished from  $|1\rangle$  and instead treat  $L_1$  as a free parameter in our simulations (see below).

The histograms of the post-selected shots in Fig. S4 demonstrate that post-selection rejects runs where leakage on those transmons occurred. Thus, while leakage may considerably impact the error-detection rate in the experiment [S11], we do not expect it to significantly affect the fidelity of the logical initialization, and gates.

## Density-matrix simulations

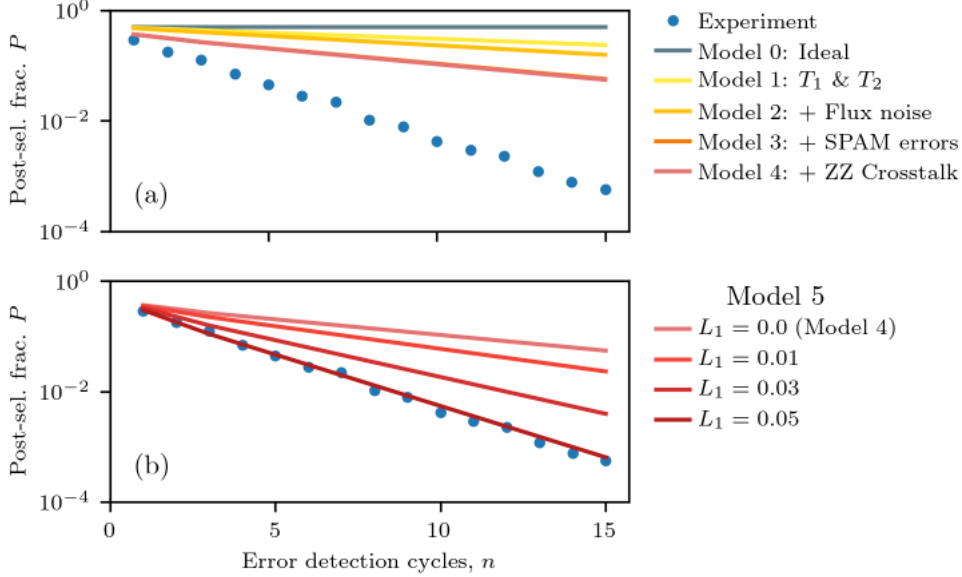


Figure S5. **Simulation of error-detection rate.** Post-selected fraction  $P$  as a function of the number  $n$  of error-detection cycles. The experimental  $P$  (blue dots) is compared to numerical simulation under various models (solid curves). (a) Simulated  $P$  obtained by incremental addition of error sources starting from the no-error (Model 0, gray); qubit relaxation and dephasing (Model 1, yellow); extra dephasing due to flux noise away from the sweetspot (Model 2, amber); state preparation and measurement errors (Model 3, orange); and crosstalk due to residual  $ZZ$  interactions (Model 4, red). (b) Simulated  $P$  for Model 5 adding CZ gate leakage with 4 different values of  $L_1$ , the leakage per CZ gate, assumed equal for all CZ gates.

We perform numerical density-matrix simulations using the *quantumsim* package [S12] to study the impact of the expected error sources on the performance of the code. We focus on repetitive error detection using the pipelined scheme and with the logical qubit initialized in  $|0_L\rangle$ . In Fig. S5a, we show the post-selected fraction  $P(n)$  as a function of the number  $n$  of error-detection cycles for a series of models. Model 0 is a no-error model, which we take as the starting point of the comparison. Model 1 adds amplitude and phase damping experienced by the transmon. Model 2 adds the increased dephasing away from the sweetspot arising from flux noise. Model 3 adds residual qubit excitation and readout (SPAM) errors. Finally, Model 4 adds crosstalk due to the residual  $ZZ$  coupling during the coherent operations of the stabilizer measurement circuits. The details of each model and their input parameters drawn from experiment are detailed below. We find that the dominant contributors to the error-detection rate are SPAM errors and decoherence. However, we also observe that the noise sources included through Model 4 clearly fail to quantitatively capture the decay of the post-selected fraction observed in experiment.

We believe that an important factor behind the observed discrepancy is the presence of leakage, as suggested by the single-shot readout histograms in Fig. S4. We consider the leakage per CZ gate  $L_1$  as a free parameter and assume the same value for all CZ gates. We simulate the post-selected fraction for a range of  $L_1$  values, shown in Fig. S5b. We observe that  $L_1 \approx 5\%$  produces a good match with experiment, suggesting that leakage significantly impacts the error-detection rate observed. This value of  $L_1$  is significantly higher than achieved in Ref. S8, which used the same device. However, note that in this earlier experiment CZ gates were characterized while keeping all other qubits in  $|0\rangle$ . Spectator transmons with residual  $ZZ$  coupling to either of the transmons involved in a CZ gate can increase  $L_1$  when not in  $|0\rangle$  (which is certainly the case in the present experiment). Note that leakage may also be further induced by the measurement [S13], an effect that we do not consider in our simulation. However, the assumption that all CZ gates have the same  $L_1$ , the approximations used in our models, and other error sources that we have not considered here may lead to an overestimation of the true  $L_1$ .

Leakage is an important error source to consider in quantum error correction experiments of larger distance codes, requiring either post-selection based on detection [S11] or the use of leakage reduction units [S14]. We leave the detailed investigation of the exact leakage rates in our experiment and the mechanisms leading to them to future work.

### Error models

Lastly, we detail the error models used in the numerical simulations in Fig. S5.

#### Model 1

We take into account transmons decoherence by including an amplitude-damping channel parameterized by the relaxation time  $T_1$  and a phase-damping channel parameterized by the pure-dephasing time at the sweetspot

$$\frac{1}{T_\phi^{\max}} = \frac{1}{T_2} - \frac{1}{2T_1},$$

where  $T_2$  is the echo dephasing time (see Table S1). The qutrit Kraus operators defining these channels are detailed in Ref. S11 and we similarly introduce these channels during idling periods and symmetrically around each single-qubit or two-qubit gate (each period lasting half the duration of the gate).

#### Model 2

We consider the pure-dephasing rate  $1/T_\phi = 2\pi\sqrt{\ln 2AD_\phi} + 1/T_\phi^{\max}$  away from the sweetspot due to the fast-frequency components of the  $1/f$  flux noise, where  $D_\phi$  is the flux sensitivity at a given qubit frequency and  $A$  is the scaling parameter for the flux-noise spectral density. We use a  $\sqrt{A} \approx 3 \mu\Phi_0$ , the average of the extracted  $\sqrt{A}$  values for  $D_3$ ,  $A_1$  and  $A_2$  obtained by fitting the measured decrease of  $T_2$  as a function of the applied flux bias, following the model described above. This allows us to estimate the dephasing time at the CZ interaction and parking frequencies, which then parameterize the applied amplitude-phase damping channel inserted during those operations [S11]. We neglect the slow-frequency components of the flux noise due to the use of sudden Net Zero pulses, which echo out this noise to first order [S7, S8].

#### Model 3

We further include state-preparation and measurement errors. We consider residual qubit excitations, where instead of the transmon being initialized in  $|0\rangle$  at the start of the experiment it is instead excited to  $|1\rangle$  with a probability  $p_e$ . We extract  $p_e$  for each transmon from a double-Gaussian fit to the histogram of the single-shot readout voltages with the transmon nominally initialized in  $|0\rangle$  [S4]. We model measurement errors via the POVM operators  $M_i = \sum_{j=0}^2 \sqrt{P(i|j)} |j\rangle\langle j|$  for  $i \in 0, 1, 2$  being the measurement outcome, while  $P(i|j)$  is the probability of measuring the qubit in state  $|i\rangle$  when having prepared state  $|j\rangle$ . We extract the probability  $P(Q = |i\rangle) = \text{Tr}(M_i^\dagger M_i \rho)$  of measuring qubit Q in state  $|i\rangle$  from simulation, where  $\rho$  is the density matrix, while application of the POVM transforms  $\rho \rightarrow M_i \rho M_i^\dagger / P(Q = |i\rangle)$ . In our simulations we condition on the detection of no error and thus we calculate  $P(Q = |0\rangle)$  and then apply  $M_0$  to the state  $\rho$ . We obtain  $P(0|j)$  for  $j \in 0, 1$  from the experimental assignment fidelity matrix [S15] (where a heralded initialization protocol was used to prepare the qubits in  $|0\rangle$  [S3]) and we assume  $P(0|2) = 0$ , consistent with the observed histograms in Fig. S4. At the end of each experiment with  $n$  error-detection cycles we calculate the probability  $P_n^f$  of obtaining trivial syndromes from the final measurements of the data qubits (see Results). From this and from the probability  $P_n(A_i = |0\rangle)$  of measuring ancilla  $A_i$  in  $|0\rangle$  at cycle  $n$ , we calculate the post-selected fraction of experiments defined as  $P(n) = P_n^f \prod_n \prod_{i=1}^3 P_n(A_i = |0\rangle)$ .

*Model 4*

We consider the crosstalk due to residual  $ZZ$  interactions between transmons. The CZ gates involved in a parity check are jointly calibrated to minimize phase errors for the whole check as one block (see Fig. S2). Instead of modeling this crosstalk as an always-on interaction and taking into account the details of the check calibration, we instead capture the net effect of this noise by including it as single-qubit and two-qubit phase errors in each CZ gate. This assumes that the crosstalk only occurs between transmons that are directly coupled, which the measured frequency shifts observed in Fig. S1 validate. We characterize the phases picked up during the CZ gates using  $k \times 2^{k-1}$  Ramsey experiments for a check involving a total of  $k$  transmons (including the ancilla). In each experiment, we perform a Ramsey experiment on one transmon labelled  $Q_k$ .  $Q_k$  is initialized in a maximal superposition using a  $R_x^{-\pi/2}$  pulse, while the remaining  $k-1$  transmons are prepared in each of the  $2^{k-1}$  computational states  $|l\rangle$ . Following this initialization, the parity check is performed, followed by a rotation of  $R_\phi^{-\pi/2}$  (while the other transmons are rotated back to  $|0\rangle$ ) and by a measurement of  $Q_k$ . By varying the axis of rotation  $\phi$ , we extract the phase  $\phi_{\text{Ram}}^k(l)$  picked up by  $Q_k$  with the remaining transmons in state  $|l\rangle$ . We perform this procedure for each of the  $k$  transmons of the check, resulting in a total of  $k \times 2^{k-1}$  measured phases, which are arranged in a column vector  $\vec{\phi}^{\text{Ram}}$ . We parameterize each CZ gate used in the parity check by a matrix  $\text{diag}(1, e^{i\phi_{01}}, e^{i\phi_{10}}, e^{i\phi_{11}})$ . The column vector  $\vec{\phi}^{\text{CZ}}$  then contains all of the phases parameterizing each of the  $k-1$  CZ gates involved in the parity checks, with  $k=3$  for the  $Z_{D1}Z_{D3}$  and  $Z_{D2}Z_{D4}$  checks and  $k=5$  for the  $Z_{D1}Z_{D2}Z_{D3}Z_{D4}$  check. We can express each of the measured phases in the Ramsey experiment as a linear combination of the acquired phases as a result of the CZ interactions between transmons, i.e.,  $\vec{\phi}^{\text{Ram}} = A\vec{\phi}^{\text{CZ}}$ , where the matrix  $A$  encodes the linear dependence. Given the measured  $\vec{\phi}^{\text{Ram}}$  we perform an optimization to find the closest  $\vec{\phi}^{\text{CZ}}$  as given by

$$\begin{aligned} \min_{\vec{\phi}^{\text{CZ}}} & \sum_i \left( \sum_j A_{ij} \vec{\phi}_j^{\text{CZ}} - \vec{\phi}_i^{\text{Ram}} \right)^2, \\ \text{subject to} & \quad 0 \leq \vec{\phi}_j^{\text{CZ}} < 2\pi. \end{aligned}$$

The optimal  $\vec{\phi}^{\text{CZ}}$  then captures the net effect of the  $ZZ$  crosstalk during the parity checks, which we include in the simulation. We do not model phase errors accrued during the ancilla readout, since in our simulation we condition on each ancilla being measured in  $|0\rangle$ .

*Model 5*

We model leakage due to CZ gates following the model and numerical implementation presented in Ref. S11. Here, we do not consider the phases picked up when non-leaked transmons interact with leaked ones (the leakage-conditional phases [S11]) and we set them to their ideal values. We also neglect higher-order leakage effects, such as excitation to higher-excited states or leakage mobility. Thus, we only consider the exchange of population between  $|11\rangle$  and  $|02\rangle$  given by  $4L_1$ , except for the CZ between  $A_1$  and  $D_3$ , where the population is instead exchanged with  $|20\rangle$  as we use the  $|11\rangle$ - $|20\rangle$  avoided crossing for this gate [S8].

There remain several relevant error sources beyond those included in our numerical simulation. For example, we do not include dephasing of data or other ancilla qubits induced by ancilla measurement, which we expect to be a relevant error source for comparing the performance of the pipelined and parallel schemes. Also, we only consider the net effect of crosstalk due to residual  $ZZ$  interactions during coherent operations of the parity-check circuits, which we include via errors in the single-qubit and two-qubit phases in the CZ gates. Thus, we do not capture the crosstalk present whenever an ancilla is projected to state  $|1\rangle$  by the readout but declared to be in  $|0\rangle$  instead. Furthermore, as  $ZZ$  crosstalk does not commute with the amplitude damping included during the execution of the circuit, we are not capturing the increased phase error rate that this leads to.

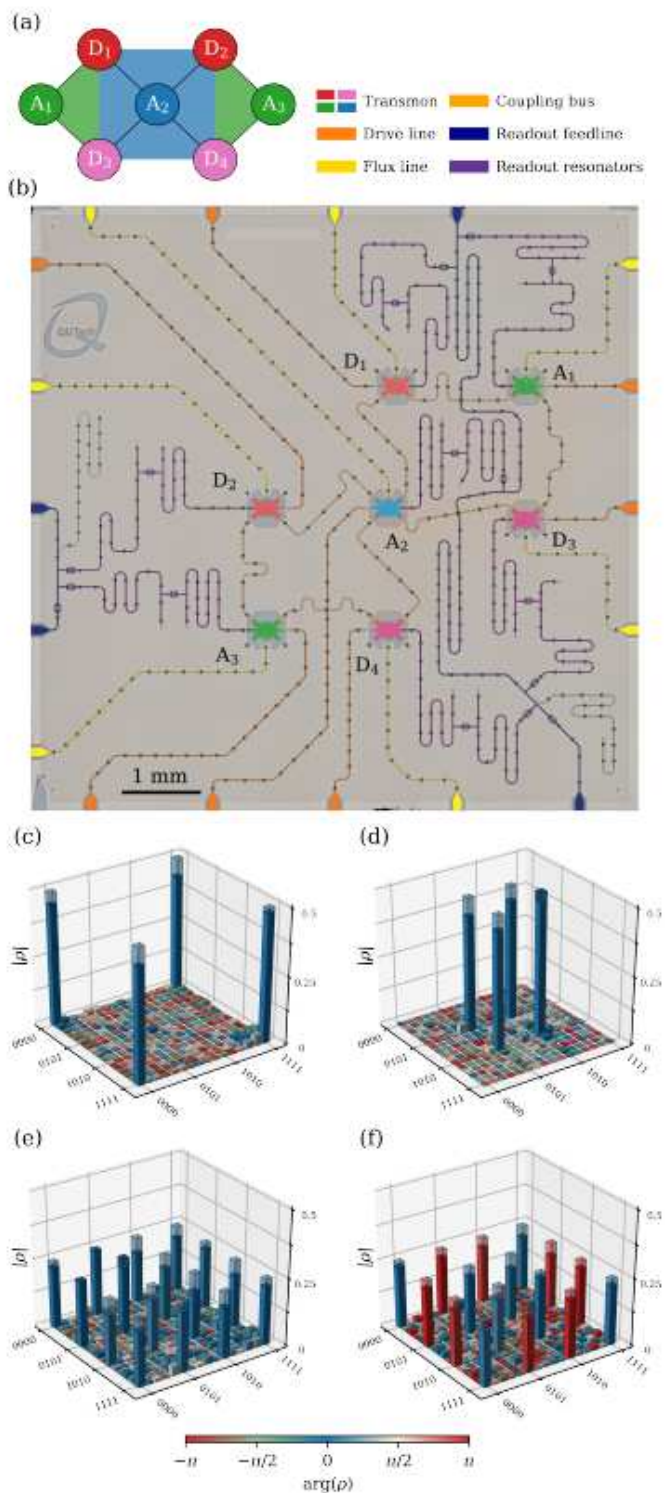
---

[S1] Krantz, P. *et al.* A quantum engineer's guide to superconducting qubits. *App. Phys. Rev.* **6**, 021318 (2019).



- [S2] Bultink, C. C. *et al.* General method for extracting the quantum efficiency of dispersive qubit readout in circuit qed. *App. Phys. Lett.* **112**, 092601 (2018).
- [S3] Ristè, D., van Leeuwen, J. G., Ku, H.-S., Lehnert, K. W. & DiCarlo, L. Initialization by measurement of a superconducting quantum bit circuit. *Phys. Rev. Lett.* **109**, 050507 (2012).
- [S4] Walter, T. *et al.* Rapid High-Fidelity Single-Shot Dispersive Readout of Superconducting Qubits. *Phys. Rev. App.* **7**, 054020 (2017).
- [S5] Schreier, J. A. *et al.* Suppressing charge noise decoherence in superconducting charge qubits. *Phys. Rev. B* **77**, 180502(R) (2008).
- [S6] Sagastizabal, R. *et al.* Variational preparation of finite-temperature states on a quantum computer. Preprint at <http://arXiv.org/abs/2012.03895> (2020).
- [S7] Rol, M. A. *et al.* Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits. *Phys. Rev. Lett.* **123**, 120502 (2019).
- [S8] Negirneac, V. *et al.* High-fidelity controlled-z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor. Preprint at <http://arXiv.org/abs/2008.07411> (2020).
- [S9] Wood, C. J. & Gambetta, J. M. Quantification and characterization of leakage errors. *Phys. Rev. A* **97**, 032306 (2018).
- [S10] Asaad, S. *et al.* Independent, extensible control of same-frequency superconducting qubits by selective broadcasting. *npj Quantum Inf.* **2**, 16029 (2016).
- [S11] Varbanov, B. M. *et al.* Leakage detection for a transmon-based surface code. *npj Quantum Information* **6**, 102 (2020).
- [S12] O'Brien, T. E., Tarasinski, B. M. & DiCarlo, L. Density-matrix simulation of small surface codes under current and projected experimental noise. *npj Quantum Information* **3** (2017).
- [S13] Sank, D. *et al.* Measurement-induced state transitions in a superconducting qubit: Beyond the rotating wave approximation. *Phys. Rev. Lett.* **117**, 190503 (2016).
- [S14] Battistel, F., Varbanov, B. M. & Terhal, B. M. A hardware-efficient leakage-reduction scheme for quantum error correction with superconducting transmon qubits. Preprint at <http://arXiv.org/abs/2102.08336> (2021).
- [S15] Heinsoo, J. *et al.* Rapid high-fidelity multiplexed readout of superconducting qubits. *Phys. Rev. App.* **10**, 034040 (2018).

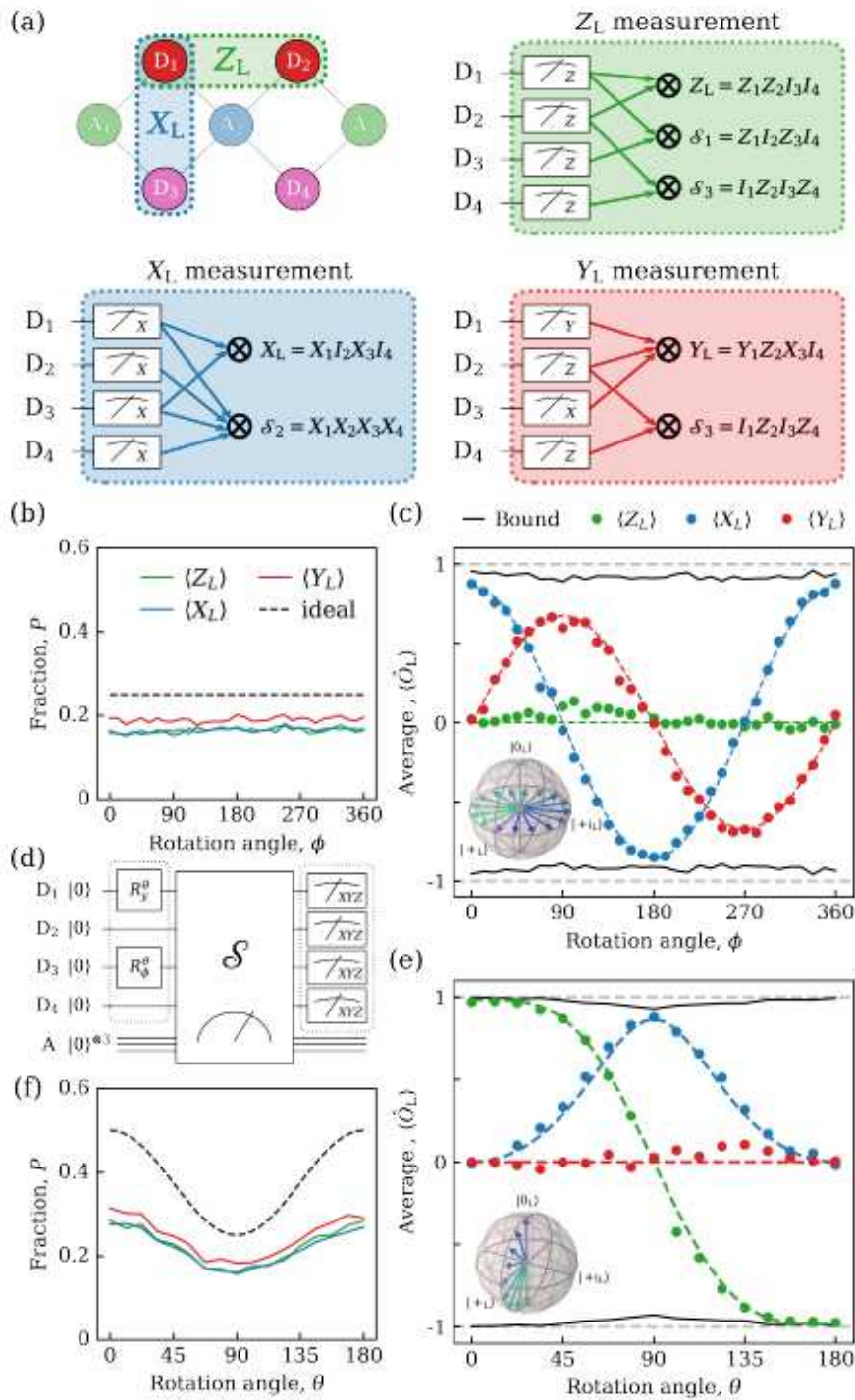
# Figures



**Figure 1**

Surface-7 quantum processor and initialization of logical cardinal states. (a) Distance-two surface code. (b) Optical image of the quantum hardware with added false-color to emphasize different circuit elements. (c-f) Estimated physical density matrices,  $\hat{\rho}$ , after targeting the preparation of the logical

cardinal states  $|0_L\rangle$  (c),  $|1_L\rangle$  (d),  $|+\rangle_L$  (e) and  $|-\rangle_L$  (f). Each state is measured after preparing the data qubits in  $|0000\rangle$ ,  $|1010\rangle$ ,  $|++++\rangle$  and  $|++--\rangle$ , respectively. The ideal target state density matrix is shown in the shaded wireframe.



**Figure 2**

Arbitrary logical-state initialization and measurement in the logical cardinal bases. (a) Assembly of data-qubit measurements used to evaluate logical operators  $Z_L$ ,  $X_L$  and  $Y_L$  with additional error detection. (d)

Initialization of logical states using the procedure described in Eq. 6. (c, e) ZL, XL and YL logical measurement results as a function of the gate angles  $\theta$  (c) and  $\phi$  (e). The colored dashed curves show a fit of the analytical prediction based on Eqs. 9 and 11 to the data and the dark curve denotes a bound based on the measured FL of each state. (b, f) Total fraction  $P$  of post-selected data as a function of the input angle for each logical measurement. The dashed curve shows the ideal fraction given by Eq. 8.

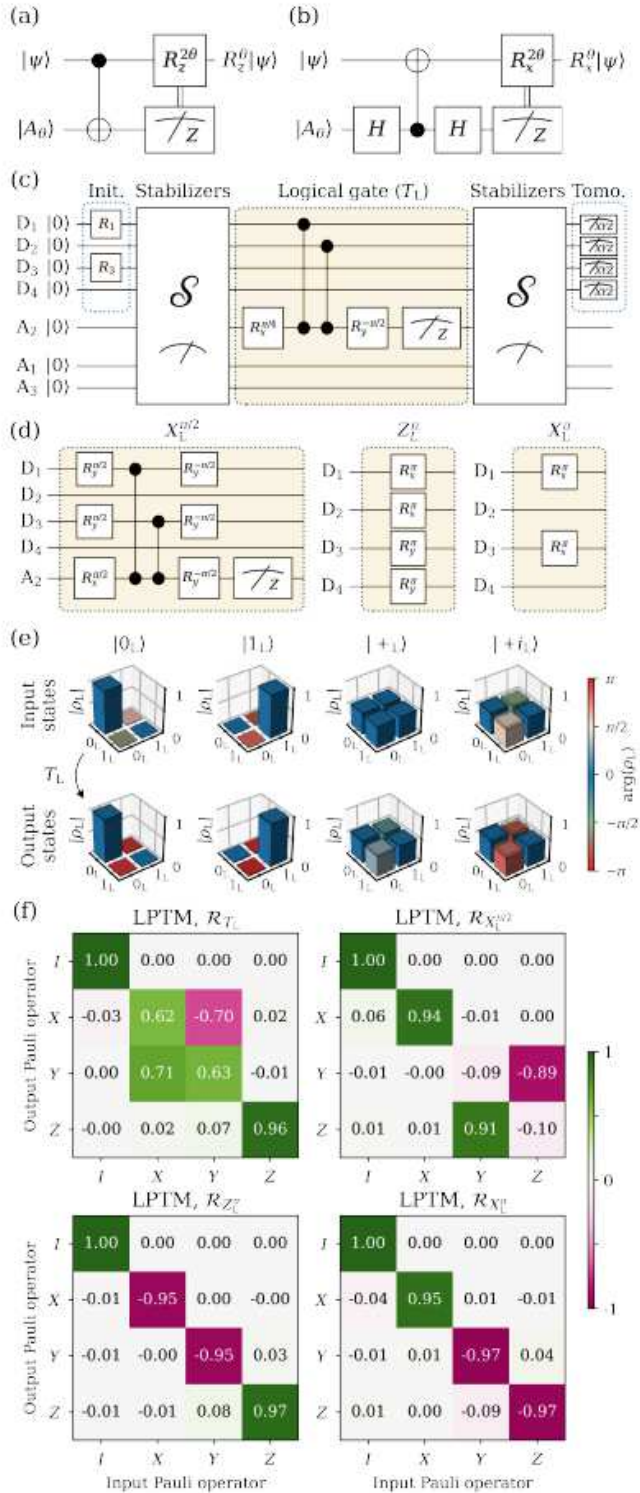
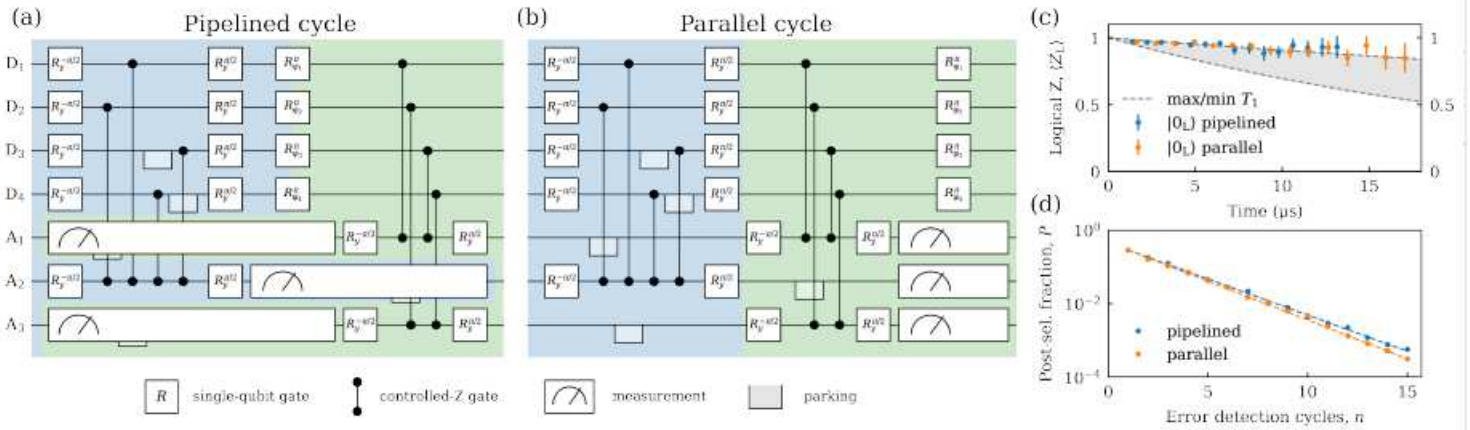


Figure 3

Logical gates and their characterization. (a, b) General gate-by-measurement schemes realizing arbitrary rotations around the Z (a) and X (b) axis of the Bloch sphere. (c) Process tomography experiment of the TL gate. Input cardinal logical states are initialized using the method of Fig. 2. Output states are measured following a second round of stabilizer measurements. (d) Logical  $X\pi/2$  L,  $Z\pi$  L and  $X\pi$  L gates compiled using our hardware-native gateset. (e) Logical state tomography of input and output states of the TL gate. These logical density matrices are obtained by performing four-qubit tomography of the data qubits and then projecting onto the codespace. (f) Extracted logical Pauli transfer matrices.



**Figure 4**

Repetitive error detection using pipelined and parallel stabilizer measurement schemes. (a, b) Gate sequences used to implement the pipelined (a) and parallel (b) stabilizer measurement schemes. Gate duration is 20 ns for single-qubit gates, 60 ns for controlled-Z (CZ) gates and parking [14, 22], and 540 ns for ancilla readout. The order of CZs in the  $XD_1XD_2XD_3XD_4$  stabilizer (blue shaded region) prevents the propagation of ancilla errors into logical qubit errors [23]. The total cycle duration for the pipelined (parallel) scheme is 840 ns (1000 ns). (c) Estimated ZL expectation value,  $\langle Z_L \rangle$ , measured for the  $|0_L\rangle$  state versus the duration of the experiment using the pipelined (blue) and the parallel (orange) schemes. We also plot the excited-state probability (right axis) set by the maximum and minimum physical qubit  $T_1$ . (d) Post-selected fraction of data versus the number of error detection cycles  $n$  for the pipelined (blue) and parallel (orange) scheme.