nature portfolio

https://doi.org/10.1038/s41586-022-04566-8

Supplementary information

Realizing repeated quantum error correction in a distance-three surface code

In the format provided by the authors and unedited

Supplementary Information
Realizing Repeated Quantum Error Correction in a Distance-Three Surface Code

Sebastian Krinner,[1,*] Nathan Lacroix,[1,*] Ants Remm,[1] Agustin Di Paolo,[2] Elie Genois,[2] Catherine Leroux,[2] Christoph Hellings,[1] Stefania Lazar,[1] Francois Swiadek,[1] Johannes Herrmann,[1] Graham J. Norris,[1] Christian Kraglund Andersen,[1,†] Markus Müller,[3,4] Alexandre Blais,[2,5] Christopher Eichler,[1] and Andreas Wallraff[1,6]

[1] *Department of Physics, ETH Zurich, CH-8093 Zurich, Switzerland*
[2] *Institut Quantique and Département de Physique, Université de Sherbrooke, Sherbrooke J1K2R1 Québec, Canada*
[3] *Institute for Quantum Information, RWTH Aachen University, Aachen D-52056, Germany*
[4] *Peter Grünberg Institute, Theoretical Nanoelectronics, Forschungszentrum Jülich, Jülich D-52425, Germany*
[5] *Canadian Institute for Advanced Research, Toronto, ON, Canada*
[6] *Quantum Center, ETH Zurich, 8093 Zurich, Switzerland*

(Dated: February 21, 2022)

CONTENTS

I. DEVICE FABRICATION

To fabricate the 17-qubit surface code device, we pattern transmon qubit islands, couplers, resonators, and control lines into a 150 nm-thin niobium film sputtered onto a high-resistivity silicon substrate using photolithography and reactive ion etching. To establish a well-connected ground plane at microwave frequencies and realize cross-overs for coplanar waveguides, we fabricate aluminum-titanium-aluminum trilayer airbridges onto the device using a two-layer resist photolithography process with reflow. We fabricate aluminium-based Josephson junctions of the transmon qubits using electron-beam lithography (EBL) and shadow evaporation, and establish electrical contact between the junction metal and the niobium base-layer film using aluminum bandages [1] fabricated in a second EBL and evaporation run.

II. DEVICE PARAMETERS AND PERFORMANCE

We characterize the performance and properties of each qubit on the device using spectroscopy and standard time-domain methods, see Tab. SI. Note that

* These authors contributed equally to this work.
† Current affiliation: QuTech and Kavli Institute for Nanoscience, Delft University of Technology, Delft 2628 CJ, Netherlands

some parameters (marked with the superscript 'a' in Tab. SI) were extracted from measurements in a separate cooldown of the same device.

We realize single-qubit gates with 40-ns-duration microwave DRAG pulses [2] with a Gaussian envelope which has a standard deviation of $\sigma = 10\,\text{ns}$ and is truncated at $\pm 2\sigma$.

We realize two-qubit gates with net-zero flux pulses [3, 4] with an average duration of 68 ns, see Section III for a detailed description of the two-qubit gates implementation. We add 15 ns buffers before and after each flux-pulse resulting in a total average two-qubit gate duration of 98 ns.

We benchmark the performance of single- and two-qubit gate operations across the entire device using randomized benchmarking [5, 6] (single-qubit gates) and interleaved randomized benchmarking [7–9] (two-qubit gates), and display the resulting errors in Fig. S1. CZ gates that are executed simultaneously in the quantum error correction cycle (see Fig. 3a and Fig. S12) are also calibrated and benchmarked simultaneously.

Single-qubit gates have a mean error of 0.09(4) %, which is in good agreement with the mean error of 0.08 % obtained from master-equation simulations taking only decoherence into account, see Section VII. Two-qubit gates have an average gate error of 1.5(1.0) %, and a best (worst) gate error of 0.6 % (5.4 %). Two-qubit gates used to realize the stabilizer $\hat{S}^{\text{X1}}$ display larger errors, which we attribute to the low coherence times of the auxiliary qubit X1, and to the interaction with a microscopic defect (Section III) with a frequency close to the idle-frequency of qubit X1.

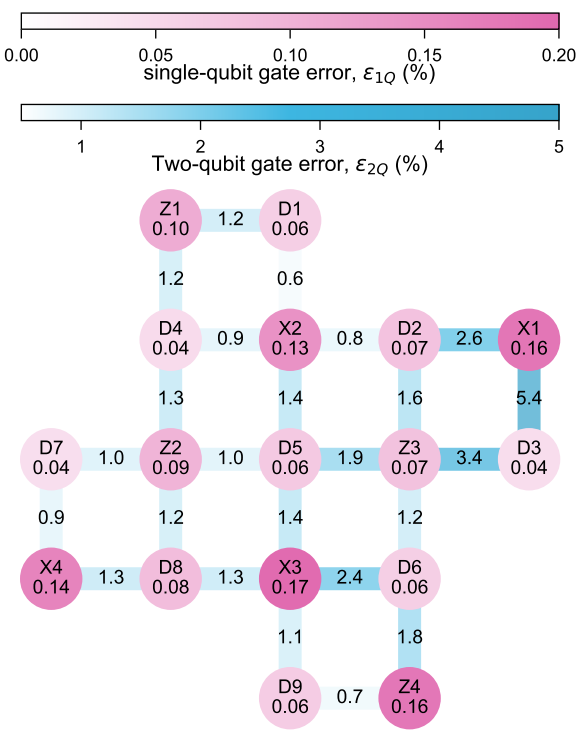

FIG. S1. Single- and two-qubit gate errors as characterized by randomized benchmarking and interleaved randomized benchmarking, respectively. The average measured single-qubit (two-qubit) gate error across the device is 0.09(4) % [1.5(1.0) %].

III. TWO-QUBIT GATES

Two-qubit gates are essential building blocks of the stabilizer circuits of the surface code. Here, we describe our implementation of CZ gates, discuss the constraints and considerations for gate parallelization and gate ordering during the error correction cycle, and report on coherence properties during the gate execution.

A. CZ gate implementation

We realize two-qubit controlled-phase (CZ) gates [10, 11] by flux-tuning two neighboring qubits to an interaction frequency between the idle frequencies of the two qubits. We use a net-zero pulse shape to protect the qubits from low-frequency flux noise and to avoid the buildup of long-timescale distortions in the flux-control lines [3, 4].

The flux pulse applied to each qubit consists of two Gaussian-filtered, approximately 30-ns-long sections, which are interleaved with a 2.5-ns-long net-zero transition section, see Fig. S2a for an example waveform. The amplitudes of the longer sections bring the $|20\rangle$-transition frequency into resonance with the the $|11\rangle$-transition frequency, see Fig. S2b. We calibrate the duration of these sections to achieve full population recovery into the $|11\rangle$ state.

To achieve a π-controlled-phase rotation, we calibrate the pulse amplitude of the transition section for the data qubit, which effectively controls the phase acquired by the $|20\rangle$ state relative to the $|11\rangle$ state.

We select interaction frequencies which allow for the parallel execution of three CZ gates in each of the eight time steps of the quantum error correction cycle, as indicated by the color and label of the qubit-qubit couplers in Fig. S2c. The interaction frequencies for a single time step (colored horizontal lines of same color) are chosen to avoid crossing neighboring qubits on which gates are executed in the same time step. In addition, the interaction frequencies in a given CZ gate time step are distributed over the available $\sim 2\,\text{GHz}$ frequency range to minimize residual-ZZ couplings between neighboring qubits performing CZ gates in parallel. The overall, calculated residual-ZZ coupling during the eight CZ gate time steps, averaged over all 24 qubit pairs and over the eight time steps, amounts to $\alpha_{\text{ZZ}}/2\pi = 27(37)\,\text{kHz}$, which is about a factor three larger than with all qubits biased at their idle frequency.

Finally, the interaction frequencies are chosen to mini-

TABLE SI. Qubit parameters, coherence properties and single-qubit performance for the nine data qubits (top) and the eight auxiliary qubits (bottom). We also provide, for relevant quantities, the averaged value across the device in column $\overline{\text{Q}}$.

Parameter	D1	D2	D3	D4	D5	D6	D7	D8	D9
Qubit idle frequency, $\omega_Q/2\pi$ (GHz)	3.885	3.994	3.952	3.878	3.895	3.74	4.056	3.993	4.143
Qubit anharmonicity, $\alpha/2\pi$ (MHz)	-184	-184	-183	-184	-186	-184	-181	-183	-181
Lifetime, T_1 (µs)	29.1	33.0	65.5	59.3	32.2	60.2	36.0	33.1	25.9
Ramsey decay time, T_2^* (µs)	35.0	13.5	33.9	74.3	46.1	78.1	74.8	41.4	31.7
Echo decay time, T_2^e (µs)	46.2	52.2	49.3	75.5	56.1	89.3	72.5	59.1	36.3
Single-qubit RB error[a], ϵ_{1Q} (%)	0.06	0.07	0.04	0.04	0.06	0.06	0.04	0.08	0.06
Readout frequency, $\omega_{RO}/2\pi$ (GHz)	6.769	6.979	6.88	7.12	7.18	7.032	6.91	7.075	6.868
Qb. freq. during RO, $\omega_Q'/2\pi$ (GHz)	5.321	4.75	5.275	4.25	4.42	5.13	4.395	3.993	5.0
Dispersive shift[a], $\chi/2\pi$ (MHz)	-2.0	-2.2	-1.9	-1.6	-1.5	-1.6	-2.3	-2.0	-2.4
Disp. shift during RO, $\chi'/2\pi$ (MHz)	-7.4	-3.9	-6.0	-2.0	-2.1	-4.7	-3.0	-2.0	-4.9
Readout linewidth[a], $\kappa_{\text{eff}}/2\pi$ (MHz)	6.3	7.2	8.3	8.7	3.5	9.0	8.6	8.2	8.9
Qubit-RO res. coupling[a], $g_{Q,RR}/2\pi$ (MHz)	244	269	241	241	238	244	267	265	260
Two-state readout error, $\epsilon_{RO}^{(2)}$ (%)	0.7	0.5	0.5	0.6	0.8	0.5	2.3	1.3	0.5
Three-state readout error, $\epsilon_{RO}^{(3)}$ (%)	4.4	1.0	5.9	2.1	1.6	1.1	3.3	2.1	1.0
Thermal population, P_{th} (%)	2.1	0.7	1.5	3.8	2.7	2.2	0.0	0.8	1.6

Parameter	X1	X2	X3	X4	Z1	Z2	Z3	Z4	$\overline{\text{Q}}$
Qubit idle frequency, $\omega_Q/2\pi$ (GHz)	6.097	5.885	6.022	6.049	6.328	6.192	5.956	6.037	-
Qubit anharmonicity, $\alpha/2\pi$ (MHz)	-170	-174	-170	-170	-163	-168	-171	-170	**-177**
Lifetime, T_1 (µs)	12.4	17.4	18.5	12.6	17.0	42.7	29.7	27.5	**32.5**
Ramsey decay time, T_2^* (µs)	5.6	14.3	20.7	28.9	29.3	33.1	48.0	28.9	**37.5**
Echo decay time, T_2^e (µs)	15.8	33.0	16.1	33.1	31.5	26.0	53.6	53.6	**47.0**
Single-qubit RB error[a], ϵ_{1Q} (%)	0.16	0.13	0.17	0.14	0.1	0.09	0.07	0.16	**0.09**
Readout frequency, $\omega_{RO/2\pi}$ (GHz)	7.372	7.554	7.258	7.461	7.316	7.502	7.2	7.412	-
Qb. freq. during RO, $\omega_Q'/2\pi$ (GHz)	5.9	5.885	6.022	6.049	6.328	6.191	5.956	5.687	-
Dispersive shift[a], $\chi/2\pi$ (MHz)	-2.8	-1.9	-3.2	-2.6	-4.7	-2.9	-3.2	-2.8	**-2.4**
Disp. shift during RO, $\chi'/2\pi$ (MHz)	-2.1	-1.9	-3.2	-2.6	-4.7	-2.9	-3.2	-2.8	**-3.5**
Readout linewidth[a], $\kappa_{\text{eff}}/2\pi$ (MHz)	15.1	20.1	13.0	11.0	11.2	12.2	14.3	10.0	**10.3**
Qubit-RO res. coupling[a], $g_{Q-RR}/2\pi$ (MHz)	167	168	167	168	171	170	167	171	**213**
Two-state readout error, $\epsilon_{RO}^{(2)}$ (%)	2.7	0.7	0.8	0.8	1.3	0.4	0.4	0.5	**0.9**
Three-state readout error, $\epsilon_{RO}^{(3)}$ (%)	3.9	2.0	1.9	1.6	2.2	1.2	0.9	1.1	**2.2**
Thermal population, P_{th} (%)	0.2	0.2	0.5	0.3	0.6	0.5	0.2	0.9	**1.1**

[a] Measured in a different cooldown.

mize population loss due to interaction with the strongly coupled microscopic defects of our device. To characterize the defect-mode distribution, we determine the frequency-dependent population loss for each qubit by measuring the remaining excited-state population after applying Gaussian-filtered square flux pulse that tunes the qubit frequency from its idle frequency (black semicircle) to ω_{int} for a duration of $t_{\text{int}} = 100\,\text{ns}$, see the gray filled areas in Fig. S2c. For most qubits, we observe a constant background population loss of $\leq 2\,\%$ over the entire frequency range, with $0-3$ narrow frequency bands ($\leq 50\,\text{MHz}$) exhibiting peak population loss $\geq 25\%$. We attribute these population-loss peaks to coherent interactions with defects coupled to the qubits with a strength of $g/2\pi \geq 0.8\,\text{MHz}$. Qubits D7, D8 and X1 display broader and higher population-loss peaks, likely due to the interaction with defects coupled to the qubits with $g/2\pi$ on the order of 1-20 MHz. For these qubits, the finite interaction during the rising and falling edge of the flux pulse leads to a population loss tail when crossing the defect. We choose interaction frequencies for the two-qubit gates that are detuned from all defects and that avoid crossing strongly coupled defects.

In addition to a suitable selection of interaction frequencies, accurate control of the qubit frequency is essential for the realization of high-fidelity and low-leakage two-qubit gates. However, the flux pulses reaching the sample and controlling the qubit frequency, are transformed from the programmed waveforms on the AWGs due to the transfer functions of, for instance, the high-pass filtering in the bias-tee, the frequency-dependent attenuation in the cables, and imperfections in the impedance matching of the flux line. To characterize the flux pulse distortions at timescales ranging from $\sim 50\,\text{ns}$ to $50\,\mu\text{s}$, we apply a step-like Gaussian-filtered flux pulse and resolve the resulting time dependence of the qubit frequency by identifying which drive frequency induces a transition from $|0\rangle$ to $|1\rangle$, for a varying delay Δt after applying the step-like flux pulse. The time-dependent qubit-frequency response is then used to calibrate infinite impulse response (IIR) filters which invert the distortions of signals propagating along the flux line in digital

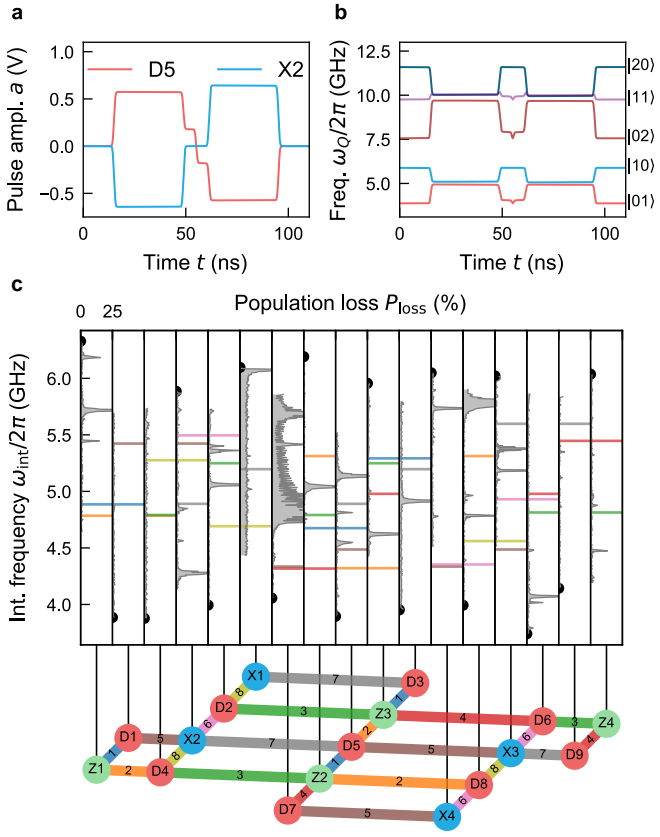

FIG. S2. Realization of two-qubit gates. **a**. Example of a net-zero pulse shapes before compensating for flux line distortions that are applied simultaneously on a data qubit (D5) and an auxiliary qubit (X2) to implement a CZ gate between the two. Note that the duration of the transition section has been increased from 2.5 ns to 12.5 ns for better visibility in this figure. **b**. Calculated evolution of the eigenenergies for the pulse shapes shown in panel **a**. **c**. Frequency arrangement chosen for the operation of the device. Measured population loss (gray areas) when flux-tuning qubits from their idle frequencies (black dots) to ω_{int} for a duration of 100 ns. The chosen two-qubit gate interaction frequencies are displayed as colored lines corresponding to the color code used in the conceptual device representation at the bottom of the panel, which indicates two-qubit pairs between which gates are executed. Gates sharing the same color are executed simultaneously, in the time step indicated by the label extending between pairs of qubits

preprocessing of the programmed waveforms, see above. Moreover, to compensate for distortions on nanosecond timescales, we calibrate finite impulse response (FIR) filters using methods described in Ref. 12.

B. Two-qubit gate order

We choose the order of the two-qubit gates, executed in the CZ gate time steps 1, 2, 3, 4 for $\hat{S}^{\text{Z}i}$ and in the time steps 5, 6, 7, 8 for $\hat{S}^{\text{X}i}$ (Fig. S2c), to provide resilience against single auxiliary-qubit errors propagating to data qubits [13, 14]. Simultaneously, we satisfy constraints imposed by the presence of microscopic defects near the two-qubit gate interaction frequencies (Section III A).

A single $\hat{X}$ error on an auxiliary qubit Zi (Xi) during a stabilizer measurement $\hat{S}^{\text{Z}i}$ ($\hat{S}^{\text{X}i}$) results in $\hat{Z}$ ($\hat{X}$) errors on all data qubits which subsequently perform CZ gates with that auxiliary qubit, see Fig. S3a for an example. Auxiliary qubit $\hat{Z}$ errors commute with the CZ gates and therefore do not propagate to data qubits. We only consider the middle of weight-four stabilizer gate sequences as potential times when $\hat{X}$ errors occur on the auxiliary qubits since the error propagates to two data qubits only in this case. For all other auxiliary qubit error locations, the error effectively propagates to at most a single data qubit because data qubit errors are only relevant up to multiplication with stabilizer operators. For instance, although an error between the first and second CZ gate physically propagates to three data qubits of the corresponding stabilizer plaquette, this three-qubit error is equivalent to a single-qubit error, of the same type, on the originally unaffected data qubit of the plaquette.

To ensure that such correlated two-qubit $\hat{Z}$ or $\hat{X}$ errors do not result in logical errors in the decoding process, we choose the last two CZ gates of each of the $\hat{S}^{\text{Z}i}$ ($\hat{S}^{\text{X}i}$) measurements to involve data qubits which are not aligned parallel to the data-qubit strings forming the logical operators $\hat{Z}_{\text{L}}$ ($\hat{X}_{\text{L}}$) [14], i.e. not horizontal (vertical) in Fig. S3b. With this gate order, correlated two-qubit $\hat{Z}$ or $\hat{X}$ errors can be correctly identified (up to multiplication with stabilizer operators). As an example, we consider the correlated error $\hat{Z}_{\text{D4}}\hat{Z}_{\text{D7}}$ on D4 and D7 resulting from a single error $\hat{Z}_{\text{Z2}}$ on Z2, see Fig. S3b. As a consequence, X2 and X4 show syndrome elements of 1 (dark blue, solid circles in Fig. S3b). Because the stabilizers $\hat{S}^{\text{X2}}$ and $\hat{S}^{\text{X4}}$ do not share a data qubit, this syndrome cannot be caused by a single data-qubit error and the minimum-weight-perfect-matching decoder (Section XII) correctly identifies both errors, the error $\hat{Z}_{\text{D7}}$ and one of the equivalent errors $\hat{Z}_{\text{D4}}$ (indicated in Fig. S3b with solid, dark blue arrows) or $\hat{Z}_{\text{D1}}$. On the other hand, if the gates Z2-D7 and Z2-D8 were successively performed in either the CZ gate time steps 1 and 2, or 3 and 4, the corresponding correlated error $\hat{Z}_{\text{D7}}\hat{Z}_{\text{D8}}$ (green rectangle with dashed outline in Fig. S3b) would not be correctly identified by the decoder. In that case, only the syndrome element of X3 is 1 and as a result the decoder identifies one of the equivalent single-qubit errors $\hat{Z}_{\text{D9}}$ (indicated in Fig. S3b with a dashed, dark blue arrow) or $\hat{Z}_{\text{D6}}$, and applies the corresponding correction $\hat{Z}_{\text{D9}}$ or $\hat{Z}_{\text{D6}}$. However, one such false correction together with the original error $\hat{Z}_{\text{D7}}\hat{Z}_{\text{D8}}$ is equivalent to the application of $\hat{Z}_{\text{L}}$ and thus leads to a logical error, see black dashed line in Fig. S3b.

We note that data qubits participating in the last two CZ gates of a stabilizer measurement may also be aligned diagonally with respect to the data qubit string of the corresponding logical operator, and that the resulting

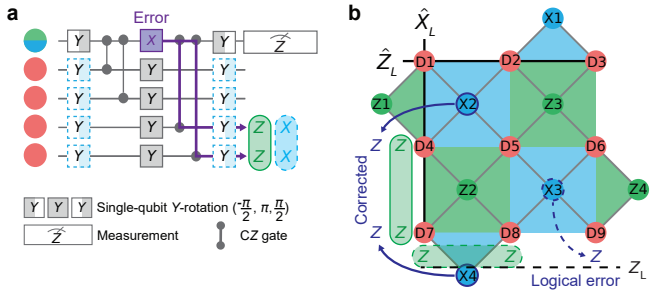

FIG. S3. Limiting the propagation of auxiliary-qubit errors. **a** Weight-four stabilizer circuit as discussed in the main text. A single $\hat{X}$ error on the auxiliary qubit in the middle of the circuit propagates to two data qubits (violet lines), resulting in a correlated $\hat{Z}\hat{Z}$ error (green box) for Z-type stabilizer circuits and in a correlated $\hat{X}\hat{X}$ error (blue rounded-corner box with dashed outline) for X-type stabilizer circuits. **b** 17-qubit surface code schematic with two potential correlated $\hat{Z}\hat{Z}$ errors (solid and dashed green boxes) indicated. Dark blue, solid and dashed circles around auxiliary qubits indicate the corresponding syndrome elements with value 1. Dark blue, solid and dashed arrows indicate the errors identified by the decoder, see text for details.

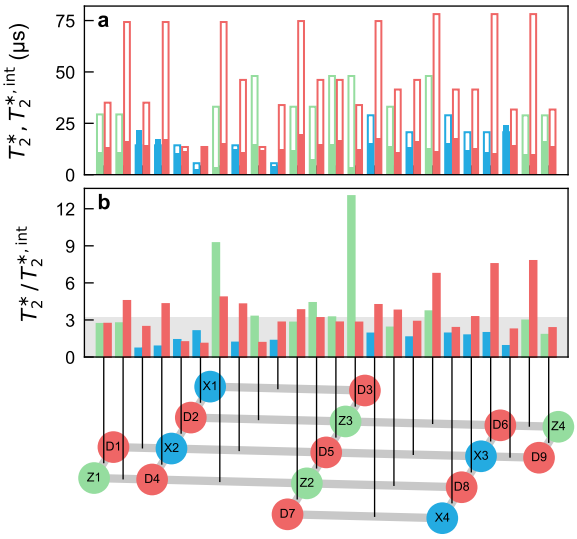

FIG. S4. Reduced coherence during two-qubit gate execution. **a** Comparison between the Ramsey decay times measured at the qubit idle frequencies (wire-frames), T_2^*, and at the two-qubit gate interaction frequencies (solid bars), $T_2^{*,\mathrm{int}}$, for the 24 two-qubit gates of the device, with colors matching the respective qubit color in the conceptual device representation at the bottom of the figure. Each qubit appears multiple times as it is tuned to different interaction frequencies depending on which neighboring qubit it performs a gate with. **b** Ratio of the Ramsey decay time at the qubit idle frequency, T_2^*, and the Ramsey decay time at the two-qubit gate interaction frequency, $T_2^{*,\mathrm{int}}$. The median ratio is indicated by the gray area.

weight-two error on the data qubits remains correctable. In particular, this is the case for $\hat{S}^{\mathrm{Z3}}$ and allows us to execute the gate Z3-D5 in a different time step than the gate Z2-D7, which would otherwise result in a frequency crossing of Z2 and D5 because the gate Z2-D7 is constrained to an interaction frequency $\omega_{\mathrm{int}}/2\pi \lesssim 4.5$ GHz due to a microscopic defect on D7 (Fig. S2c). For completeness, we note that there is no constraint imposed on the two-qubit gate order from maintaining the commutation of neighboring stabilizers $\hat{S}^{\mathrm{Z}i}$ and $\hat{S}^{\mathrm{X}i}$ [14, 15] because we measure $\hat{S}^{\mathrm{Z}i}$ and $\hat{S}^{\mathrm{X}i}$ in a pipelined approach [16].

C. Reduced coherence at interaction frequencies

At the two-qubit gate interaction frequency, the sensitivity of the qubit frequency to fluctuations in flux is higher than at the idle frequency, which is first-order insensitive to fluctuations in flux. Hence, flux-noise-induced dephasing is increased during two-qubit gates. We characterize the effective Ramsey decay time $T_2^{*,\mathrm{int}}$, for a given qubit Qi ($Q \in \{\mathrm{D}, \mathrm{X}, \mathrm{Z}\}$ and $i = 1, \ldots, 9$ if $Q = \mathrm{D}$ and $i = 1, \ldots, 4$ otherwise), fluxed-tuned to the interaction frequency ω_{int} with a sequence of Ramsey measurements. Specifically, we measure the phase of a given qubit with a Ramsey experiment in which we insert a train of N 60 ns-duration net-zero flux pulses (with 15-ns-long buffers on each side) between two Ramsey $\pi/2$-pulses. We repeat the measurement for $N = 1, 2, 4, 8, \ldots, 256$ pulses, extract the phase contrast for each measurement from a cosine fit, and fit the phase contrasts to a decaying exponential to extract $T_2^{*,\mathrm{int}}$.

We compare $T_2^{*,\mathrm{int}}$ to the Ramsey decay time measured at the qubit idle frequency, T_2^*, for each of the 24 two-qubit gates, see the wire-frame and filled bars in Fig. S4a, respectively. Averaged over the 24 qubit-pairs, we observe a mean $T_2^{*\mathrm{int}}$ of 13.1(4.2) µs at the interaction frequency, compared to a mean T_2^* of 46.8(20.2) µs at the idle frequency. As the built-in echo effect of the net-zero flux-pulse provides protection against low-frequency flux-noise, we hypothesize that the reduction in decay-time is dominated by high-frequency flux noise. We compute the decay-time ratio $T_2^*/T_2^{*,\mathrm{int}}$ for each qubit-pair, see Fig. S4b, and extract a mean ratio of ~ 3.2 (gray area). This reduction in Ramsey decay time significantly affects the coherence limit of the two-qubit gates. To reproduce the characteristics of the experiment in numerical simulations, we account for this reduction by adjusting the dephasing rates for the duration of the two-qubit gates, see Section VII for details.

IV. EXPERIMENTAL SETUP

We install the 17-qubit quantum device in a magnetically-shielded sample holder mounted at the base

plate of a dilution refrigerator [17] and connect it to the control-electronics setup located at room temperature as indicated in Fig. S5. Input and output lines for charge control (pink), flux control (green), and readout (purple), are configured with the indicated microwave components for signal conditioning.

DC voltage sources (Bat.) generate a current which passes through a series of attenuators and filters to induce a magnetic flux in the SQUID-loop of the transmon qubit and control its idle frequency. Arbitrary waveform generators (AWG) generate voltage pulses at a sampling rate of 2.4 GSa/s to control the qubit frequencies on the nano-second timescale to implement two-qubit gates, see Section III. The AWG signal is combined with the DC bias current using a bias-tee.

Single-qubit drive pulses are generated at an intermediate frequency in the range of 0-500 MHz by an AWG, and then up-converted (UC) to microwave frequencies in an analog IQ-mixer using the continuous-wave signal of a microwave generator (MWG) as a carrier.

An ultra-high frequency quantum analyzer (UHFQA) generates multiplexed-readout pulses at a sampling rate of 1.8 GSa/s. The amplification chain at each readout port of the sample consists of a wide-bandwidth near-quantum limited traveling-wave parametric amplifier (TWPA) [18], a high-electron-mobility transistor (HEMT) amplifier and low-noise, and room-temperature amplifiers (RT-A board). We add the TWPA pump tone to the input of the amplifier using a 20 dB directional coupler (Dir. coupler), and cancel it interferometrically at the input of the room-temperature amplifiers by combining the cryostat output signal with a phase- and amplitude-displaced pump tone that bypasses the cryostat. After amplification, the signal is down-converted with an IQ-mixer in a down-conversion (DC) board and then both digitally demodulated and integrated using the field-programmable gate-array of the UHFQA.

V. CROSSTALK CHARACTERIZATION AND COMPENSATION

An important requirement for scaling up quantum processors is the individual and independent control of its constituents. Here, we characterize three types of crosstalk relevant for the execution of the surface code on our 17-qubit device: microwave drive crosstalk, flux crosstalk, and measurement-induced dephasing.

A. Drive crosstalk

During the execution of multi-qubit quantum circuits, individual qubits are susceptible to off-resonant driving mediated by microwave pulses applied to drive lines designed to address other qubits. Ensuring low cross-driving is therefore essential for the successful execution of the surface code.

We characterize the coupling rates of each drive line (DL Qi) to all other qubits on the device, by measuring the excited state population of qubit Qj after applying a 100-ns-long Gaussian drive pulse to drive line DL Qi at the frequency of qubit Qj. We sweep the pulse amplitude and fit a cosine function to the measured data to obtain the amplitude $a_j^{\mathrm{DL}\,Qi}$ corresponding to a π-rotation induced on qubit Qj. As a measure of cross-driving, we normalize each value by the π-pulse amplitude of the target qubit $a_i^{\mathrm{DL}\,Qi}$ to obtain the cross-driving ratios, of which we show the negative logarithm $-\log_{10}\left(a_j^{\mathrm{DL}\,Qi}/a_i^{\mathrm{DL}\,Qi}\right)$ in Fig. S6. Given the chosen pulse duration and the maximum drive rate which we can apply without compressing the up-conversion IQ-mixer, this method allows us to resolve cross-driving ratios down to about $-\log_{10}\left(a_j^{\mathrm{DL}\,Qi}/a_i^{\mathrm{DL}\,Qi}\right) \approx -1.6$, or equivalently a Rabi rotation angle of 4.6 degrees. We find the off-diagonal elements of the cross-driving matrix on average 1.5 orders of magnitude smaller than the drive rate of the target qubit, which provides an upper bound for the cross-driving of Qj via DL Qi, since pulses applied to DL Qi during the execution of the surface code are detuned from the frequency of Qj.

Some auxiliary-data qubit pairs in physical proximity exhibit larger crosstalk, most likely mediated by the qubit-qubit coupling resonator. However, for all these pairs, the drive line addresses a qubit in the other frequency band (see Fig. 1a and Fig. 1c for a visual representation of the geometry and the frequency bands) than the cross-coupled qubit (see hatched region in Fig. S6), which strongly suppresses the effective cross-driving during device operation. Taking both the relative detunings and the experimentally characterized cross-driving ratios into account, we estimate the gate error induced on Qj resulting from applying a π-pulse on Qi via DL Qi and find that the gate error is smaller than 0.01% for all auxiliary-data qubit pairs except for DL X3 - D9 and DL X2 - D1 with estimated gate errors of 0.06% and 0.17%, respectively.

For qubit pairs within the same frequency band, we find that the expected gate error is smaller than 0.2% for all pairs except for the drive line DL D5 and qubits D1 and D4, for which the expected gate errors are 19.2% and 16.6%, respectively. We attribute the stronger cross-coupling for these elements to the crossover between DL D5 and the qubit-qubit couplers connecting Z1-D1 and D4-X2 (see Fig. 1b). As crosstalk on that scale would significantly impact the execution of the surface code, we compensate the cross-driving of DL D5 on D1 and D4 with an interferometric cancellation drive pulse which has opposite amplitude but is otherwise identical to the pulse arriving at D1 via DL D5.

Note that this characterization measurement was performed in a separate cooldown of the same device.

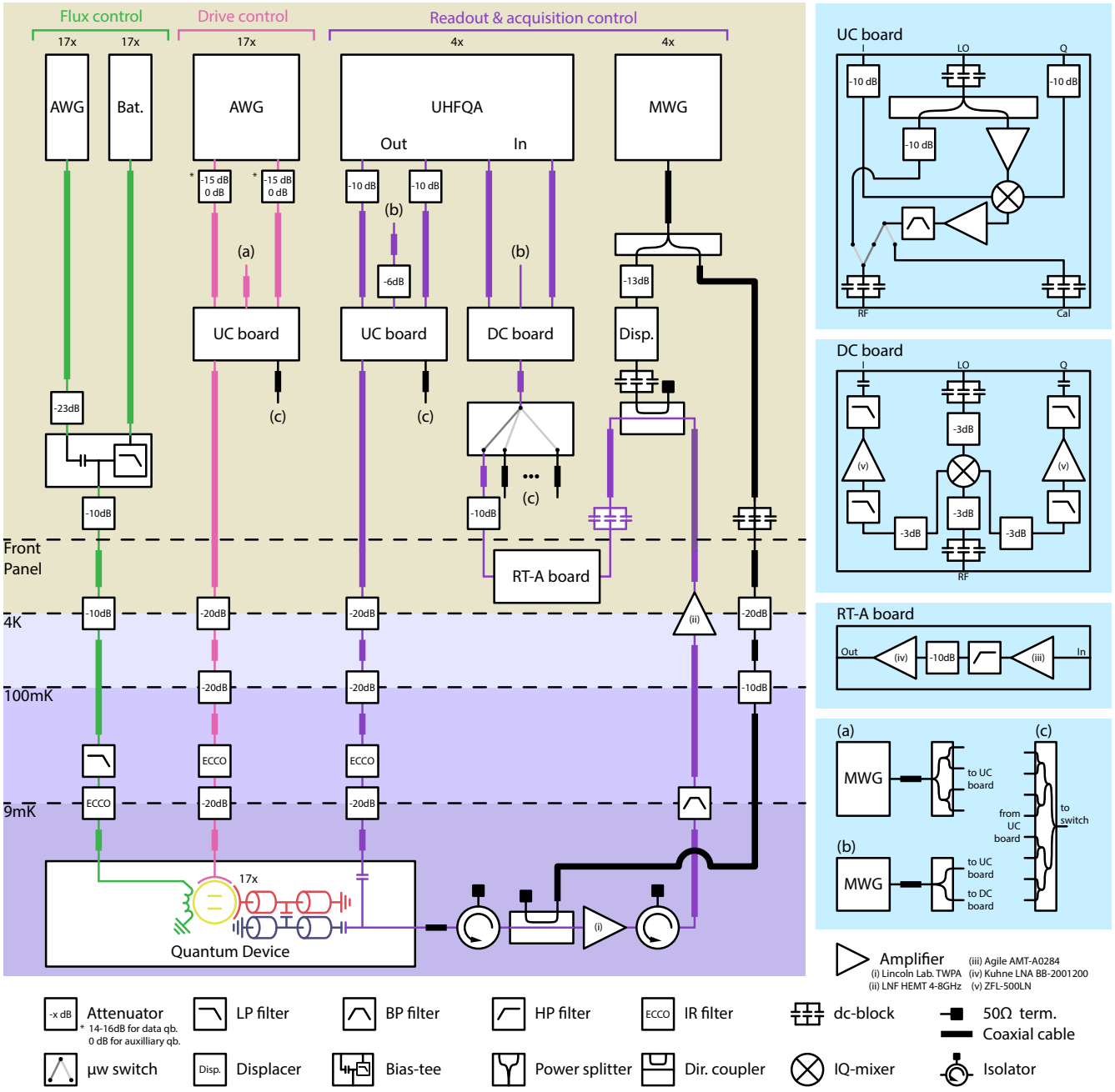

FIG. S5. Schematic illustration of the experimental setup. Flux lines (green), drive lines (pink), and readout lines (purple) connect room-temperature electronics to the device, schematically represented by qubits (yellow), readout resonators (red) and Purcell filters (dark blue). The background colors indicate the temperature stages of the experimental setup. We provide additional details about the components in the legend (white background) and in the text.

B. Flux crosstalk

During the execution of a two-qubit gate, the transition-frequencies of both qubits are tuned to an intermediate interaction frequency (Section III), at which the sensitivity of the qubit-frequency to flux is increased. Consequently, the parallel execution of two-qubit gates requires a careful flux crosstalk characterization and compensation.

We characterize the coupling of each flux line on the device to all qubits using a sequence of Ramsey experiments. In particular, we measure the effect of the flux line with target qubit Qi (FL Qi) on any qubit Qj by sweeping the amplitude V_i of a voltage pulse on FL Qi and measuring the phase ϕ_j induced on Qj in a Ramsey experiment. During the experiment, Qj is flux-tuned away from its idle frequency to increase its flux sensitiv-

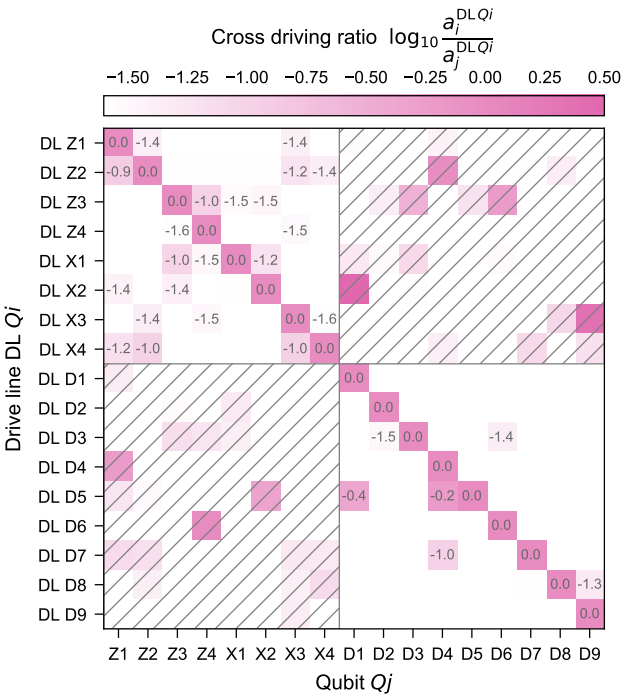

FIG. S6. Drive crosstalk characterization. For each drive line (DL) targeting qubit Qi, crosstalk is expressed as the logarithm of the ratio between the amplitude of a π-pulse on the target qubit, $a_i^{\mathrm{DL}Qi}$, and the amplitude of a π-pulse on any other qubit j, $a_j^{\mathrm{DL}Qi}$. The hatched gray area corresponds to cross-driving of far detuned qubits ($\sim$2 GHz), see text for details. White matrix elements without annotation correspond to DL-qubit pairs with cross-driving ratios < -1.6.

ity and thereby increase the signal-to-noise ratio of the measured phase. We convert the measured phase to frequency of Qj, $\omega_j = \phi_j/\tau_{\mathrm{fp}}$, with $\tau_{\mathrm{fp}} = 60\,\mathrm{ns}$ the duration of the flux pulse applied on FL Qi, which is close to the mean two-qubit gate duration on our device. Finally, we convert the frequency $\omega_j(V_i)$ to a flux $\Phi_j(V_i)$ dependent on the applied voltage V_i and fit a linear function to it, whose slope $\mathrm{d}\Phi_j/\mathrm{d}V_i$ corresponds to an element of the flux crosstalk matrix $\mathbf{C}$. Repeating this procedure for all flux lines and all qubits, we obtain the flux crosstalk matrix. For each flux line FL Qi, we normalize all crosstalk elements with respect to the targeted coupling on Qi, i.e. $\mathrm{d}\Phi_i/\mathrm{d}V_i$, to obtain the cross-flux ratio, which we show on a base-ten logarithmic scale in Fig. S7a.

We find that crosstalk (off-diagonal elements) is on average about three orders of magnitude smaller than the target couplings (diagonal elements). The total induced flux $\vec{\Phi} = (\Phi_1, \ldots, \Phi_{17})^{\mathrm{T}}$ in the SQUID loops when applying voltage pulses with amplitude $\vec{V} = (V_1, \ldots, V_{17})^{\mathrm{T}}$ is given by

$$\vec{\Phi} = \mathbf{C}\,\vec{V} \tag{1}$$

Hence, to induce a target flux vector $\vec{\Phi}'$, we program the amplitudes

$$\vec{V}' = \mathbf{C}^{-1}\,\vec{\Phi}' \tag{2}$$

which compensates for the flux line crosstalk between all qubits on the device.

To verify the effectiveness of our compensation scheme, we repeat the crosstalk characterization measurement with activated flux crosstalk compensation, which yields $\tilde{\mathbf{C}}$, see Fig. S7b. The resulting off-diagonal elements of $\tilde{\mathbf{C}}$ are suppressed by two additional orders of magnitude. To further reduce the flux crosstalk in our experiments, we apply the crosstalk compensation method recursively by using $\tilde{\mathbf{C}}^{-1}$ as a second compensation matrix.

C. Measurement-induced dephasing

A successful implementation of the surface code requires repeated readout of a subset of auxiliary qubits without disturbing the state of any other qubits. However, in the presence of finite readout crosstalk, a readout pulse on an auxiliary qubit can off-resonantly excite the readout resonator of another qubit, thereby leading to dephasing and/or coherent phase rotations on that other qubit.

To characterize this effect, we sweep the amplitude of the readout pulse of an auxiliary qubit Qi while measuring the phase of another qubit Qj in a Ramsey experiment [19]. We fit the Ramsey-fringes contrast to a Gaussian model and the phase deviation to a quadratic model, from which we extract the measurement-induced dephasing rate Γ_{ij}, and the coherent phase rotation $\Delta\phi_{ij}$ for the untargeted qubit Qj when reading out Qi. The additional dephasing rate can be related to a phase-flip probability $P_\phi^{ij} = [1-\exp(-\Gamma_{ij}\tau_{\mathrm{RO}})]/2$, where τ_{RO} is the readout pulse duration.

We characterize the measurement-induced dephasing of all auxiliary qubits on any other qubit, see Fig. S8. Note that for this measurement, the qubit idle frequency of X1 was changed from 6.097 GHz to 4.429 GHz to avoid a microscopic two-level defect whose frequency drifted towards X1's idle frequency.

On average, we observe a phase-flip probability of 0.09 % and a coherent phase rotation of 0.6°. In our implementation of the surface code, all Z-type (X-type) auxiliary qubits are measured simultaneously, and therefore their mutual measurement-induced dephasing is not a concern (hatched region in Fig. S8). We observe the largest dephasing ($P_\phi = 2.3\,\%$ and $\Delta\phi = 13.4°$) on auxiliary qubit Z2, when reading out auxiliary qubit X2, which we attribute to cross-driving of the readout resonator of Z2 by the readout signal targeting X2. This dephasing could partially explain the higher mean syndrome element σ_m^{Z2}, compared to the other mean syndrome elements (see Fig. 3c).

In future work, we expect that we could correct for the coherent phase rotations by utilizing virtual-Z rotations of equal magnitude and opposite sign.

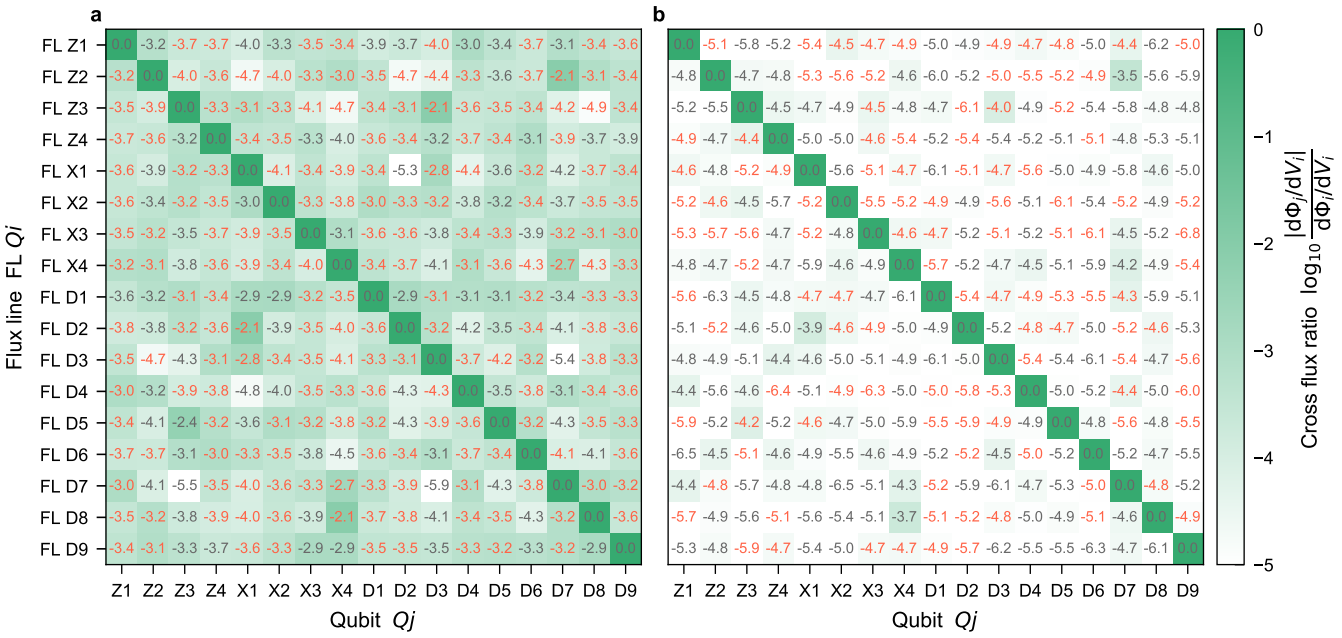

FIG. S7. Flux crosstalk characterization and compensation. **a** Measured flux crosstalk matrix **C** normalized by its diagonal, for a 60 ns net-zero flux pulse and without crosstalk compensation (see text for details). **b** Measured flux crosstalk matrix with a single round of flux crosstalk compensation calibrated based on the matrix shown in **a**. Red numbers indicate a negative sign of $\mathrm{d}\Phi_j/\mathrm{d}V_i$.

VI. READOUT CHARACTERIZATION

A. Three-state readout

We dispersively read out the state of the transmon qubits by applying a Gaussian-filtered ($\sigma = 10$ ns) rectangular pulse with a pulse duration of 200 ns to auxiliary qubits and 300 ns to data qubits, see Fig. S10b for an example waveform. To optimally distinguish between the first three states of the transmon, we integrate the complex-valued downconverted signals $s(t)$ in real-time on the UHFQA (see Section IV) with two sets of complex-valued, 400-ns-long integration weights $w_i(t)$ [19–23] to obtain

$$u_i = \mathrm{Re}\left\{\int_0^{t_{\mathrm{int.}}} s(t) w_i(t)\,\mathrm{d}t\right\}. \tag{3}$$

We use the integration weights

$$w_1(t) = s^*_{|1\rangle}(t) - s^*_{|0\rangle}(t), \tag{4}$$

$$w_2(t) = s^*_{|2\rangle}(t) - s^*_{|0\rangle}(t) - \tag{5}$$

$$- \frac{\int w_1(t)(s_{|2\rangle}(t) - s_{|0\rangle}(t))\,\mathrm{d}t}{\int |w_1(t)|^2\,\mathrm{d}t} w_1(t), \tag{6}$$

where $s_i(t)$ is the averaged measured readout-resonator response for a qubit prepared in state $i \in \{|0\rangle, |1\rangle, |2\rangle\}$.

To characterize the single-shot readout, we prepare the qubit 10^5 times in each of the three basis states ($|0\rangle$, $|1\rangle$, and $|2\rangle$), and measure the integrated resonator response yielding a pair of values $\{u_1, u_2\}$ for each experimental run. Prior to applying the state-preparation pulses we perform a pre-selection readout and reject measurements in which the qubit is not in the ground state. We estimate the thermal population for each qubit as the probability to be in $|1\rangle$ or $|2\rangle$ after this pre-selection readout, see Tab. SI. We fit the distribution of measured $\{u_1, u_2\}$ pairs to a trimodal Gaussian mixture model, associating each qubit state to one of the Gaussian components, see Fig. S9. To characterize two-state readout, we use only the Gaussian components of the mixture model corresponding to the $|0\rangle$ and $|1\rangle$ state. Based on the fitted model, we assign the individual readout outcomes to the most likely qubit state and compute the N-level readout error (Tab. SI)

$$\epsilon_{\mathrm{RO}}^{(N)} = 1 - \frac{1}{N}\sum_{i=1}^{N} P(i|i), \tag{7}$$

where $P(i|i)$ is the probability of correctly assigning the state i to a qubit prepared in the state i.

B. Flux pulse-assisted readout

We make use of the flux-tunability of our transmon qubits to dynamically change the qubit frequency for the duration of the readout, see Fig. S10a. This allows us to optimize readout parameters *in-situ*, such as the readout resonator-qubit detuning and the dispersive shift χ,

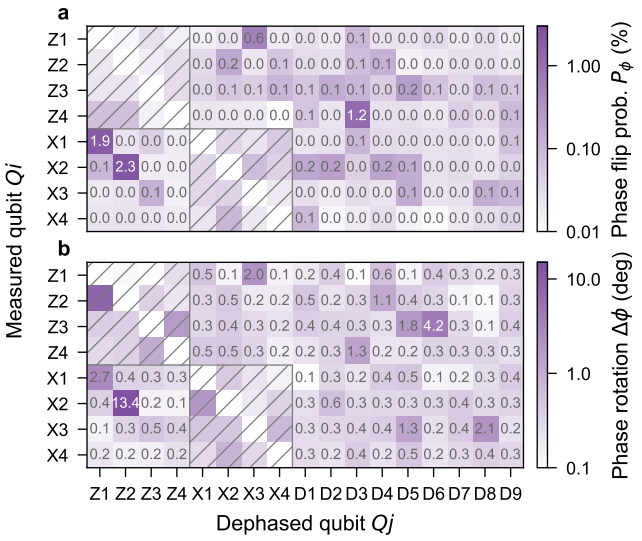

FIG. S8. Measurement-induced dephasing. **a** Phase flip probability and **b** coherent phase rotation when measuring auxiliary qubits, on any other qubits. The hatched gray area corresponds to auxiliary qubits which are read out simultaneously in the surface-code experiment (see text for details).

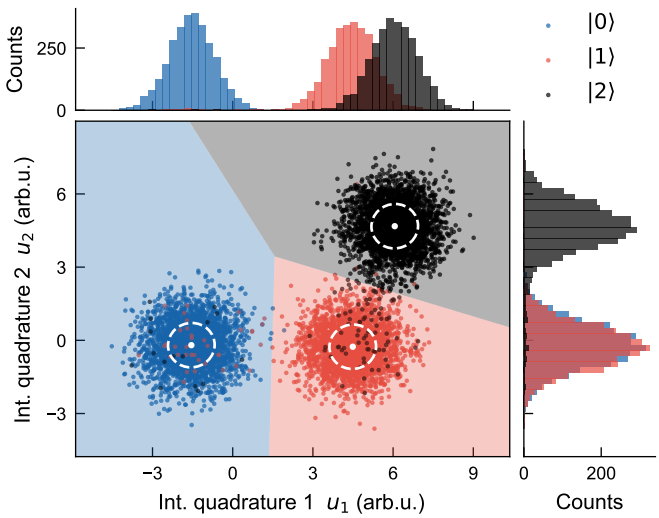

FIG. S9. Three-state readout characterization of auxiliary qubit X4 presented as a representative example. We show the first 3000 of the 10^5 single-shot measurements for each of the three state preparations, after preselection. The blue, red, and black areas delimit the regions of the integrated quadrature plane in which a measured data point is assigned to $|0\rangle$, $|1\rangle$ and $|2\rangle$, respectively. We indicate the mean (white dot) and the 1σ confidence ellipse (white dashed line) of each Gaussian distribution. The marginal histogram distributions of u_1 and u_2 are shown in the top and right panel, respectively.

see Tab. SI for the parameters during readout. We employ this method for all data qubits to reduce the detuning with the readout resonator during the readout, and for auxiliary qubit X1 to avoid a microscopic defect located close to its idle frequency (Fig. S2c).

We use Gaussian-filtered, rectangular flux pulses with short rising and falling edges ($\sigma = 0.5$ ns) in order to minimize coupling to defects. The flux pulse lasts longer than the readout pulse (see also Section X) because we continue integrating the readout signal while the readout resonator field is ringing down.

VII. NUMERICAL SIMULATIONS

A. General considerations

In this section, we describe how the simulations of the experiment are performed. These simulations are based on the measured device characteristics including the individual qubit coherence times, readout errors (see Tab. SI) and spurious ZZ interactions.

For the simulation of two-qubit gates we include the increased dephasing rates at the two-qubit gate interaction frequencies according to the measurements described in Section III C. We note that the simulations provide an upper bound for the performance achievable with the specified device parameters because further error sources such as gate control errors, population loss into microscopic defect modes and measurement-induced dephasing are not included in the simulation model. We observed that by modeling the two-qubit gate operation with an increased qubit dephasing rate reproducing the gate fidelities measured in randomized benchmarking, we find logical lifetimes in the simulation which match closely the experimental values. We opted against including these dephasing rates into the simulations, to avoid overinterpreting the predictive power of the model.

1. *Device Hamiltonian*

The system Hamiltonian includes a flux-tunable transmon Hamiltonian per qubit mode and capacitive-coupling interactions of the form $\hbar g_{k,k'} \hat{n}_k \hat{n}_{k'}$, where $\hat{n}_k$ is the charge operator associated with the kth qubit, and $\hbar g_{k,k'}$ is the capacitive coupling strength between the pair of qubits (k, k'). Single-qubit gates are modeled by a microwave-drive Hamiltonian of the form $2e\hat{n}_k V_k(t)$, where $V_k(t)$ is a time-dependent voltage. Two-qubit gates are modeled by time-dependent flux pulses $\{\phi_k(t), \phi_{k'}(t)\}$ on both target and control qubits in the pair (k, k'). Here, ϕ_k is the external flux applied to the SQUID loop of qubit k.

To facilitate the comparison to spectroscopy data used for gate calibration, we move to the flux-dependent basis which diagonalizes the device Hamiltonian, resulting in

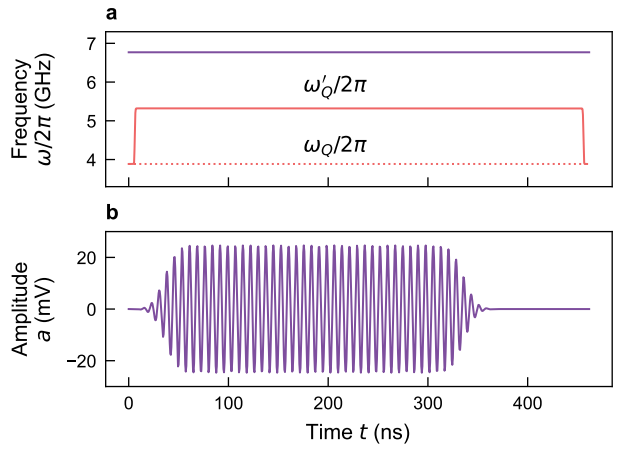

FIG. S10. Pulse sequence used for flux pulse-assisted readout of qubit D1. **a** The qubit transition-frequency (red solid line) is tuned from its lower sweet spot value (red dotted line), $\omega_Q \sim 4\,\mathrm{GHz}$, to a value ω'_Q closer to the readout resonator frequency (purple line), using a fast flux pulse. **b** Intermediate frequency signal programmed on the AWG to generate the readout pulse.

an effective model of the form

$$\begin{aligned}\hat{H}_{\mathrm{eff}}(t)/\hbar = &\sum_k \omega_k[\phi_k(t)]b_k^\dagger b_k + \frac{\alpha_k}{2}b_k^{\dagger 2}b_k^2 - i\Omega_k(t)(b_k - b_k^\dagger)\\ &+ \sum_{k<k'}\xi_{k,k'}[\phi_k(t),\phi_{k'}(t)]b_k^\dagger b_k b_{k'}^\dagger b_{k'}\\ &+ \dots\end{aligned} \tag{8}$$

Here, $\omega_k[\phi_k(t)]$, α_k and $\Omega_k(t)$ correspond, respectively, to the frequency, anharmonicity and effective drive amplitude of the kth qubit mode, while $\xi_{k,k'}[\phi_k(t),\phi_{k'}(t)]$ is the strength of the cross-Kerr (or ZZ) interaction of the form $b_k^\dagger b_k b_{k'}^\dagger b_{k'}$ between the pair of modes (k,k'). Note that the explicit dependence of the qubit frequencies and cross-Kerr interactions on the flux biases ϕ_k is omitted below. This model accounts for the change in the ZZ-interactions with the bare mode frequencies, and therefore with the external flux biases $\{\phi_k(t),\phi_{k'}(t)\}$ during the two-qubit gates. Higher-order interactions represented by the dots in Eq. (8) are not included in this effective model. Indeed, the impact of such nonidealities is minimized by device design and careful scheduling of the two-qubit gates, as described in Sections III and V.

We describe relaxation and dephasing with the effective zero-temperature master equation for the system's density matrix $\rho(t)$

$$\begin{aligned}\dot{\rho}(t) = &-i[\hat{H}_{\mathrm{eff}}(t)/\hbar,\rho(t)]\\ &+\sum_k \gamma_{1,k}\mathcal{D}[b_k]\rho(t)\\ &+\sum_k \gamma_{\varphi,k}[\phi_k(t)]\mathcal{D}[b_k^\dagger b_k]\rho(t),\end{aligned} \tag{9}$$

where $\gamma\mathcal{D}[\hat{o}]\bullet = \mathcal{D}[\hat{c}]\bullet = \hat{c}\bullet\hat{c}^\dagger - \{\hat{c}^\dagger\hat{c},\bullet\}/2$ for the collapse operator $\hat{c} = \sqrt{\gamma}\hat{o}$, and $\gamma_{1,k}$ and $\gamma_{\varphi,k}[\phi_k(t)]$ are, respectively, the decay and dephasing rates associated to mode k. The dephasing rates of the qubits involved in two-qubit gates incorporate a flux-bias dependence $\gamma_{\varphi,k}[\phi_k(t)]$ which is determined experimentally.

2. *Numerical solver*

To make the simulation of the 17-qubit chip numerically tractable, with each qubit modeled as a d-dimensional Kerr-nonlinear oscillator, we employ the method of Monte Carlo wavefunctions [24–26] as implemented in QuTiP's `mcsolve` [27]. Because it evolves wavefunctions of size d^{17} rather than the $d^{17}\times d^{17}$ density matrix necessary for a master-equation simulation, the memory requirements are significantly reduced in this approach with respect to the master-equation simulation.

Succinctly, the Monte Carlo method evolves a stochastic wavefunction $|\psi(t)\rangle$ according to a non-Hermitian Hamiltonian $H_{\mathrm{nh}} = H_{\mathrm{eff}}(t) - i\frac{\hbar}{2}\sum_l c_l^\dagger c_l$, where c_l are the collapse operators of the system's master equation. Time-evolution under H_{nh} for a time dt leads to a decrease of the norm-square of the wavefunction by $dp = \sum_l dt\langle\psi(t)|c_l^\dagger c_l|\psi(t)\rangle \ll 1$. At time $t+dt$, the wavefunction $|\psi(t+dt)\rangle$ is renormalized with probability $1-dp$, or subject to a single quantum jump with probability dp. In the case of a jump, the wavefunction collapses to the state $c_l|\psi(t)\rangle/\langle\psi(t)|c_l^\dagger c_l|\psi(t)\rangle^{1/2}$ with relative probability $\langle\psi(t)|c_l^\dagger c_l|\psi(t)\rangle/\sum_{l'}\langle\psi(t)|c_{l'}^\dagger c_{l'}|\psi(t)\rangle$.

One realization of this stochastic evolution is known as a quantum trajectory, and an advantage of this approach is that multiple such trajectories can be numerically computed in parallel [25]. With a sufficiently large number of trajectories, one can recover the solution of the master equation in Eq. (9) as

$$\mathbb{E}[|\psi(t)\rangle\langle\psi(t)|] = \rho(t). \tag{10}$$

Expectation values of an observable $\hat{O}$ are similarly obtained from

$$\mathbb{E}[\langle\psi(t)|\hat{O}|\psi(t)\rangle] = \mathrm{tr}[\rho(t)\hat{O}]. \tag{11}$$

We adjust the solver's error-tolerance parameters such that Eq. (10) holds numerically (see `qutip.Options`). To do so, we determine the appropriate solver parameters by comparing the result of `qutip.mcsolve` to that produced by the complete master-equation solver `qutip.mesolve` for systems of up to seven qubits and error-tolerance parameters set to numerical accuracy. In our stochastic simulations, we determine the number of trajectories that are needed by analyzing the convergence of Eq. (11) for all the stabilizer and logical-qubit operators. We found that 50k trajectories were sufficient to estimate these expectation values with a sampling uncertainty of less than 1 %, for the reduced model that we introduce below.

3. Measurement model

Qubit measurements are modeled by letting the qubit idle for a time equal to the experimental readout time, followed by the projection of the stochastic wavefunction according to

$$|\psi\rangle \to \frac{\Pi_k^j |\psi\rangle}{\sqrt{p_k^j}}, \tag{12}$$

where $\Pi_k^j = |j\rangle\langle j|_k$ is the single-qubit projector of the qubit k, $j \in \{0,1\}$ corresponds to the measured qubit state, and $p_k^j = \langle\psi|\Pi_k^j|\psi\rangle$ is the probability of measuring qubit k in the state j. We model qubit readout errors by flipping the result of the measurement with a probability that is computed from the readout assignment matrix for each qubit.

4. Single-qubit gate model

Single-qubit gates on qubit k are implemented with a Gaussian DRAG waveform with carrier frequency ω_k [2] and additional virtual-Z gates [28] when needed. To speed up the simulations, we drop counter-rotating terms following the usual rotating-wave approximation (RWA). We test this approximation by first comparing the average gate error per qubit with and without the counter-rotating terms, which we find to be limited by decoherence instead. Second, we have found no significant discrepancy in the estimated value of the logical qubit operators when using a RWA for simulating circuits with up to seven-qubits in a setup similar to that of Ref. [29] for up to five quantum-error-detection cycles.

5. Two-qubit gate model

CZ gates between a pair of qubits (k, k') are emulated by the free evolution of the effective Hamiltonian $(\pi/t_g - \xi_{k,k'})b_k^\dagger b_k b_{k'}^\dagger b_{k'}$, where t_g is the experimental gate time. During the gate time, we change the dephasing rates $\gamma_{\varphi,k}$ and $\gamma_{\varphi,k'}$ of the two qubits involved in the gate to the effective values corresponding to the coherence times measured experimentally for the 24 different controlled-Z gates at the qubit interaction frequencies, see Section III.

Given that the qubits are flux-biased at specific interaction frequencies during the gate time, we also account for the residual cross-Kerr interactions by adjusting the interaction strengths $\xi_{k,k'}$ of the gate qubits with their neighbors on the device. With this model, we simulate an average gate infidelity of 0.9 %, which is consistent with the median of the interleaved randomized benchmarking error of 1.2 %. Data post-selection in the experiment mitigates the impact of leakage to a large extent and enables the use of a simpler model of the two-qubit gates which involves levels only within the computational subspace. As discussed in Section III, the performance of some of the two-qubit gates is limited by the interaction of qubits with strongly coupled two-level defects, an effect which is not included in the model used for the simulations.

6. Pulse schedule

Combining our model for the single- and two-qubit gates with projective measurements, we concatenate these operations to compose the pulse sequences used in the quantum error correction experiment, including buffer times, state-preparation and measurement pulses.

B. Reduced 9+4 model

By using the Monte Carlo solver we can simulate a 17-qubit system in a reasonable time using a general-purpose workstation: approximately 25 s per trajectory per error-correction cycle with a clock speed of ~ 3 GHz and using about 6 GB of RAM. However, improvements in runtime and potential extensions of this method to even larger systems are possible with effective models with a reduced number of qubits. In this section, we describe an approximate model that employs a total of 13 qubits and is obtained by tracing out the auxiliary-qubit modes that do not participate in a given stabilizer measurement.

In practice, this effective model significantly reduces the simulation requirement to about 2 s per trajectory per QEC cycle while using less than 1 GB of memory. A schematic illustration of the two models in consideration is provided in Fig. S11.

To introduce the "9+4" model, we first consider the Z-type stabilizer measurements which constitute the first half of the QEC cycle. As illustrated in Fig. S11a (see also Fig. 2a), the circuit only involves single- and two-qubit operations on the nine data and four Z-type auxiliary qubits. Meanwhile, the measurement of the X-type auxiliary qubits, which occurs in 400 ns, projects those qubits into a product state of the form $|X_n\rangle = \bigotimes_{k'' \in \mathrm{X-aux.}} |j_{k''}\rangle$ with $j_{k''} \in \{0,1\}$. By assuming that the state of the X-type auxiliary qubits remains close to a product state during the full semi-cycle duration, it is possible to trace-out those qubit modes from the Hamiltonian. Following the measurement, the best description of the auxiliary modes is given by a product of single-qubit density matrices of the form $|1_{k''}\rangle\langle 1_{k''}|e^{-t/T_{1,k''}} + |0_{k''}\rangle\langle 0_{k''}|(1 - e^{-t/T_{1,k''}})$ if the measurement result for qubit k'' is 1, and $|0_{k''}\rangle 0_{k''}|$ otherwise. In other words,

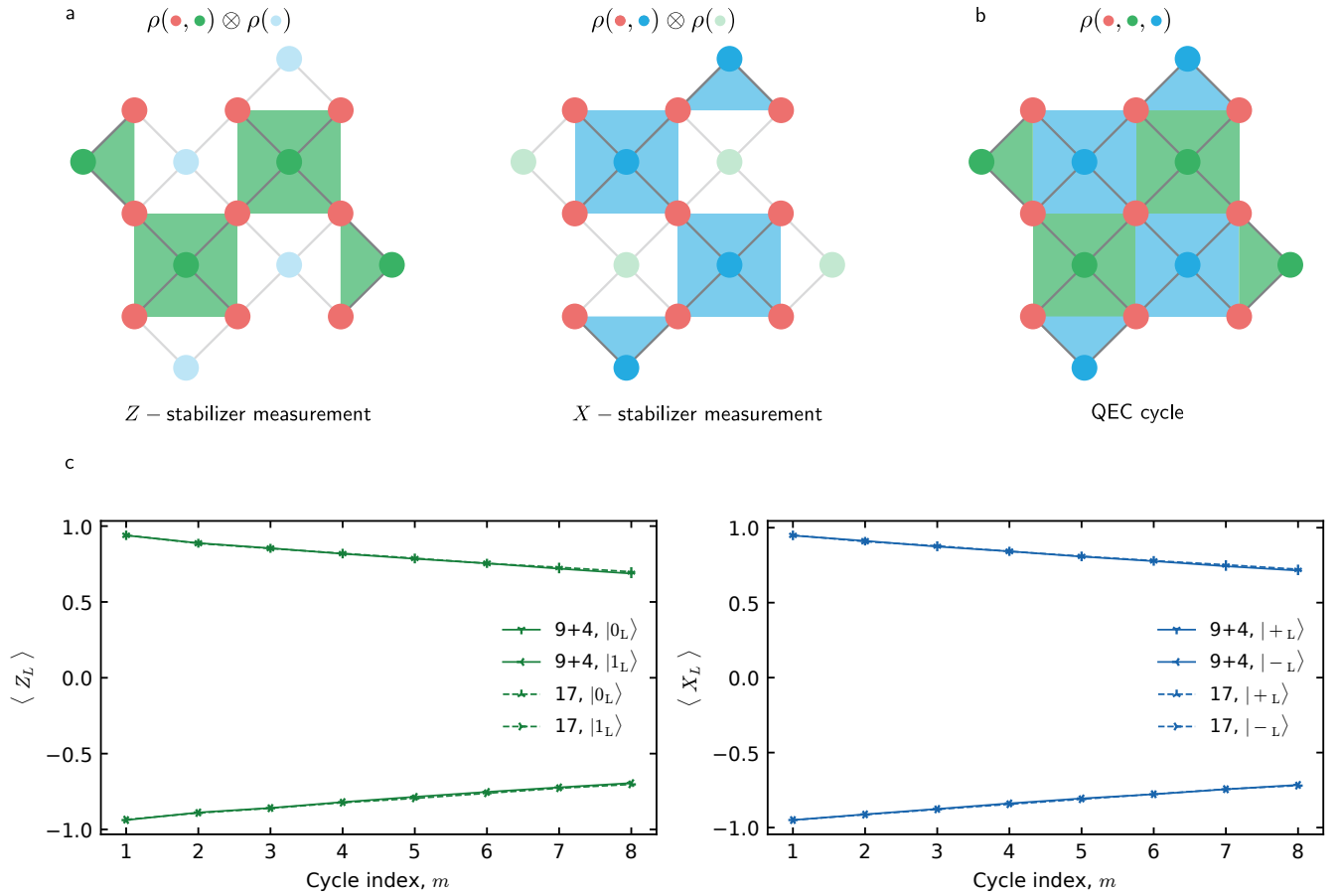

FIG. S11. **a** Illustration of the 9+4 model. Logical Z (X) stabilizer measurements are simulated by tracing out the state of the X-type (Z-type) auxiliary qubits prior to the stabilizer measurement. $\rho(\{\bullet\})$ refers to a density matrix describing the subsystem $\{\bullet\}$ of qubits indexed by color. **b** Full 17-qubit model includes all qubits at all times in the simulation. **c** Comparison between the decoded simulated data for the reduced 9+4- and the full 17-qubit model. The parameters of the simulation are those measured on the device and provided in Tab. SI. We observe an excellent agreement of the resulting logical T_1 and T_2 times between these two models.

this corresponds to the effective model

$$\begin{aligned}\hat{H}_{\rm red}(t)/\hbar = &\sum_k \Big(\omega_k + \sum_{k''\in X} \xi_{k,k''} e^{-t/T_{1,k''}} \langle b^\dagger_{k''} b_{k''}\rangle\Big) b^\dagger_k b_k \\ &+ \frac{\alpha_k}{2} b^{\dagger 2}_k b^2_k - i\Omega_k(t)(b_k - b^\dagger_k) \\ &+ \sum_{k<k'} \xi_{k,k'} b^\dagger_k b_k b^\dagger_{k'} b_{k'},\end{aligned} \tag{13}$$

where the subindices k and k' run over data and Z-type auxiliary qubits, and $\langle b^\dagger_{k''} b_{k''}\rangle$ is the expectation value of $b^\dagger_{k''} b_{k''}$ immediately after the preceding X-type auxiliary measurement. After evolving the 13-qubit wavefunction under $\hat{H}_{\rm red}(t)$ for the full semi-cycle, the projective measurement of the four Z-type auxiliary qubits is performed and the state of the system at time T is updated as

$$|\psi(T)\rangle|Z_n\rangle \to |\psi(T)\rangle|X_n\rangle, \tag{14}$$

where $|\psi(T)\rangle$ describes the state of the data qubits after the projection. We perform the simulation of the following X-type stabilizer measurement in a similar way.

In our simulations, we account for some of the noise associated with the auxiliary qubits that were traced out. First, as Eq. (13) shows, the cross-Kerr interaction between an auxiliary qubit that has been traced out and its neighboring data qubits varies in time to account for the auxiliary qubit's finite T_1 time. Second, the idling state of a traced-out auxiliary qubit k'' after its measurement is modeled by flipping $|j_{k''}\rangle$ from $|1\rangle$ to $|0\rangle$ with probability $1-\exp\left(-T_{\rm semi-cycle}/T_{1,k''}\right)$, where $T_{\rm semi-cycle}$ is the time of a quantum error correction semi-cycle. The resulting auxiliary-qubit state from this process is used to simulate the next quantum error correction cycle.

To test the validity of this reduced model, Fig. S11c shows a comparison between the results obtained for the reduced 9+4- and 17-qubit models for up to eight QEC cycles. The parameters of the simulation reflect those of the measured device, as discussed above. We show the

TABLE SII. Stabilizers $\hat{S}^{Ai}$ of the distance-3 surface code, characterized by the measured stabilizer error (ϵ) and the stabilizer error ($\epsilon_{\rm sim}$) determined from simulation.

Symbol	Stabilizer	ϵ (%)	$\epsilon_{\rm sim}$ (%)
$\hat{S}^{\rm Z1}$	$\hat{Z}_1\hat{Z}_4$	2.9	2.8
$\hat{S}^{\rm Z2}$	$\hat{Z}_4\hat{Z}_5\hat{Z}_7\hat{Z}_8$	8.4	5.0
$\hat{S}^{\rm Z3}$	$\hat{Z}_2\hat{Z}_3\hat{Z}_5\hat{Z}_6$	6.8	4.3
$\hat{S}^{\rm Z4}$	$\hat{Z}_6\hat{Z}_9$	2.5	2.0
$\hat{S}^{\rm X1}$	$\hat{X}_2\hat{X}_3$	5.7	6.7
$\hat{S}^{\rm X2}$	$\hat{X}_1\hat{X}_2\hat{X}_4\hat{X}_5$	5.9	3.9
$\hat{S}^{\rm X3}$	$\hat{X}_5\hat{X}_6\hat{X}_8\hat{X}_9$	11.8	4.4
$\hat{S}^{\rm X4}$	$\hat{X}_7\hat{X}_8$	4.5	2.6
	Weight-two average	3.9	3.5
	Weight-four average	8.2	4.4
	Average	6.1	3.9

decoded data from a logical-state preservation simulation as a function of the number of QEC cycles, from where we can infer the logical T_1 and T_2 for the two models. We use the decoder described in Section XII with weights that are extracted from the simulation data for each model. The result shows an excellent agreement between the reduced and full models of the experiment, with a discrepancy that is contained in the error of the exponential fit. This observation gives us confidence about the accuracy of the 9+4 model used to perform the simulations reported in the main text. We note that, while limited in some respects, our effective model goes beyond previous approaches [30, 31] in the treatment of correlated errors (spurious ZZ) and dissipation in continuous time.

VIII. STABILIZERS

The distance-three surface code comprises eight stabilizer operators, i.e. products of $\hat{Z}$ or $\hat{X}$ Pauli operators of a subset of the nine data qubits, as listed in Tab. SII. We characterize the quantum circuit realizing each $\hat{S}^{Ai}$ (Fig. 2a,b) for the 2^N input states of the $N \in \{2, 4\}$ data qubits of a given plaquette (Fig. 1a) and compute the stabilizer error

$$\epsilon = 1 - \frac{1}{2^N}\sum_{n=1}^{2^N}\frac{1}{2}|\overline{s}_n^{Ai} - \overline{s}_{n,\rm ideal}^{Ai}|, \tag{15}$$

see Tab. SII. We measure an average weight-two stabilizer error of 3.9 %, which is in good agreement with the 3.5 % average weight-two stabilizer error extracted from numerical simulations (Section VII). The average weight-four stabilizer error obtained from measurements (8.2 %) is larger than the one obtained from simulation (4.4 %), which we identify as most likely due to the interaction with microscopic defects, residual coherent errors in two-qubit gates, and residual crosstalk which are not modeled in our numerical simulations (Section VII).

IX. LOGICAL STATE INITIALIZATION AND CHARACTERIZATION

Here, we describe the characterization of the prepared nine-data-qubit logical states by measuring their quantum state fidelity with respect to (i) the target state, (ii) the target logical subspace, and (iii) states in correctable subspaces.

A. Fidelity with respect to the target state

While the complete tomographic reconstruction of a generic quantum state ρ of $n = 9$ qubits would require the measurement of $4^n = 262144$ independent Pauli correlators [32], the measurement of the fidelity with respect to a target logical state $\rho_{|0\rangle_{\rm L}} = |0\rangle_{\rm L}\langle 0|_{\rm L} = \hat{\mathcal{P}}_{\rm O}\hat{\mathcal{P}}_{\rm L}$ expressed in terms of the projector $\hat{\mathcal{P}}_{\rm L}$ onto the logical subspace and the projector $\hat{\mathcal{P}}_{\rm O}$ onto the $+1$ eigenstate of $\hat{Z}_{\rm L}$

$$\hat{\mathcal{P}}_{\rm O}\hat{\mathcal{P}}_{\rm L} = \frac{1}{2^9}(1+\hat{Z}_{\rm L})\prod_{i=1}^{4}(1+s^{{\rm X}i}\hat{S}^{{\rm X}i})\prod_{i=1}^{4}(1+\hat{S}^{{\rm Z}i}) \tag{16}$$

$$= \frac{1}{512}\sum_{j=1}^{512}\gamma_j\hat{P}_j \tag{17}$$

requires the measurement of only $2^n = 512$ terms [33, 34]. Here, $s^{{\rm X}i}$ are the outcomes of the X-stabilizer measurements obtained in the state initialization, $\hat{P}_i$ are 9-qubit Pauli correlators obtained from expanding the product in Eq. (16), and the γ_j take values $+1$ or -1 depending on the individual outcomes of $\{s^{{\rm X}i}\}$. The fidelity $F_{\rm phys}$ of state ρ with respect to $|0\rangle_{\rm L}$ is given by

$$F_{\rm phys} = {\rm Tr}\left(\rho\rho_{|0\rangle_{\rm L}}\right) = \frac{1}{512}\sum_{j=1}^{512}\langle\gamma_j\hat{P}_j\rangle_\rho \tag{18}$$

where $\langle\gamma_j\hat{P}_j\rangle_\rho$ corresponds to the expectation value of the Pauli correlator $\hat{P}_i$ in state ρ.

We evaluate each of the 512 Pauli correlators of Eq. (18) with data collected by executing a single quantum error correction cycle, followed by single-qubit tomography rotations on data qubits and a readout of all qubits. We observe that in 80.5% of the experimental runs the measurements of s^{Zi} yield the target value $+1$ for all four $\hat{S}^{{\rm Z}i}$, which is in good agreement with the probability of having none of the four Z-type mean syndrome elements signaling an error in the first cycle, $\prod_i^4(1-\overline{\sigma}_1^{Zi}) \approx 84.5\%$, see Fig. 3c.

For each $\langle\gamma_j\hat{P}_j\rangle_\rho$, we reject leakage events as detected by our three-state readout scheme (Section XI) and account for readout errors (Section VI) on each of the data qubits involved in the correlator. Note that apart from leakage rejection we keep all instances for further data analysis, including those events in which at least one

of the four measured $\hat{S}^{\mathrm{Z}i}$ values yields -1. Following Eq. (18), we then take the average over all expectation values to compute F_{phys}.

B. Fidelity in the logical subspace

Similarly, we evaluate the probability of having prepared a state in the logical subspace as

$$P_{\mathrm{L}} = \mathrm{Tr}\left(\rho\hat{\mathcal{P}}_{\mathrm{L}}\right) = \frac{1}{256}\sum_{j=1}^{256}\langle\gamma_j\hat{P}_j\rangle_\rho \tag{19}$$

where the sum extends over the expectation values of the 256 Pauli correlators resulting from the expansion of the projector onto the logical subspace $\hat{\mathcal{P}}_{\mathrm{L}}$. Together with the value of F_{phys} the probability P_{L} yields an estimate for the logical fidelity $F_{\mathrm{L}} = F_{\mathrm{phys}}/P_{\mathrm{L}}$, which corresponds to the fidelity of the prepared state with the target logical state conditioned on having successfully projected onto the logical subspace.

C. Fidelity with respect to states in correctable subspaces

In the context of quantum error correction, it is insightful to compute the fidelity of the prepared state with respect to any of the states which are equivalent to the target state up to a correctable error. By construction, a data-qubit state projected to the logical Z-basis (i. e. the simultaneous eigenbasis of all stabilizer operators and the logical-Z operator) is a ±1 eigenstate of the eight stabilizers $\hat{S}^{Ai}$ and a ±1 eigenstate of $\hat{Z}_{\mathrm{L}}$. Thus, there are $2^9 = 512$ distinct eigenstates of the data-qubit space in the logical Z-basis, with pair-wise degenerate syndromes. However, in a simple error correction scheme, each of the $2^8 = 256$ possible syndromes produced by the eight stabilizers is associated to a single corrective action. Consequently, for each of the 256 pairs of eigenstates with a degenerate syndrome, only one of the two eigenstates can be included in a given correctable subspace. To perform error correction most effectively, one includes the eigenstate in the correctable subspace, which can be reached by a Pauli error that is more likely to occur on the physical device.

Specifically, we construct correctable subspaces for $|0\rangle_{\mathrm{L}}$ in the following way: for each pair of eigenstates with degenerate syndrome, we compute the lowest weight errors which, when applied to $|0\rangle_{\mathrm{L}}$, lead to these two eigenstates, respectively. With the assumption that lower-weight errors are more likely than higher-weight errors, we preferably include the eigenstate that can be reached via the lowest weight error. If the two eigenstates can only be reached by applying errors of equal weight-n, we check how many different errors of weight-n lead to each of the two eigenstates, and include the eigenstate that can be reached by more weight-n errors. Finally, if both eigenstates can be reached by an equal number of weight n errors, we randomly select one of the two eigenstates. Note that with this approach, any correctable subspace includes $|0\rangle_{\mathrm{L}}$ (trivial $+1$ eigenstate of all stabilizers and of $\hat{Z}_{\mathrm{L}}$), as well as all the states, which can be reached by applying a weight-one Pauli error on any one of the data qubits.

We define the fidelity of the prepared state ρ with respect to a set of N states $\{\hat{E}_n|0\rangle_{\mathrm{L}}\}$ as

$$\begin{aligned} F_{\mathrm{c}} &= \sum_{n=1}^{N}\mathrm{Tr}\left(\rho\,\hat{E}_n|0\rangle\langle 0|_{\mathrm{L}}\hat{E}_n^\dagger\right) \\ &= \frac{1}{512}\sum_{n=1}^{N}\mathrm{Tr}\left(\rho\sum_{j=1}^{512}\hat{E}_n\gamma_j\hat{P}_j\hat{E}_n^\dagger\right) \end{aligned}$$

where we have made use of Eq. (17). We observe that $\hat{E}_n\hat{P}_j\hat{E}_n^\dagger = c_{j,n}\hat{P}_j$ with $c_{j,n} = \pm1$ because $\hat{E}_n$ and $\hat{P}_j$ are both tensors of Pauli matrices, and all matrices in the Pauli group commute or anti-commute. Consequently,

$$F_{\mathrm{c}} = \frac{1}{512}\sum_{n=1}^{N}\sum_{j=1}^{512}c_{j,n}\langle\gamma_j\hat{P}_j\rangle_\rho \tag{20}$$

can be computed based on the same set of 512 measured expectation values $\langle\gamma_j\hat{P}_j\rangle_\rho$ with the appropriate combination of sign prefactors $c_{j,n}$. This circumvents the need to measure the expectation value of 512 Pauli correlators for each of the $N = 256$ states spanning the correctable subspace.

We evaluate the fidelity F_{c} of the prepared state with respect to any state in the correctable subspace for 500 different correctable subspaces randomly chosen according to the procedure described above. We obtain a mean correctable fidelity averaged over the 500 correctable subspaces of $F_{\mathrm{c}} = 96.0(9)\,\%$. The 4.0(9)% average infidelity, which can also be interpreted as the probability of being in a state which would result in a logical error after a single cycle, is in good agreement with the probability of logical error per cycle $\epsilon_{\mathrm{L}} = 0.032(1)$ deduced from the state preservation of $\langle\hat{Z}_{\mathrm{L}}\rangle$, see main text. We observe that ϵ_{L} is slightly smaller, likely due to the fact that the state initialization characterization cannot correct for auxiliary-qubit readout errors, which require several quantum error-correction cycles to be detected.

X. PULSE SEQUENCE

We realize the quantum circuit presented in Fig. 3 a with a combination of microwave and flux pulses applied to the 17 qubits, see Fig. S12 for the complete pulse sequence of a single cycle of quantum error correction used to prepare $|0\rangle_{\mathrm{L}}$.

The sequence starts with a multiplexed readout of all qubits to herald the ground-state in post-selection. Next, single-qubit drive pulses are applied to a subset of the

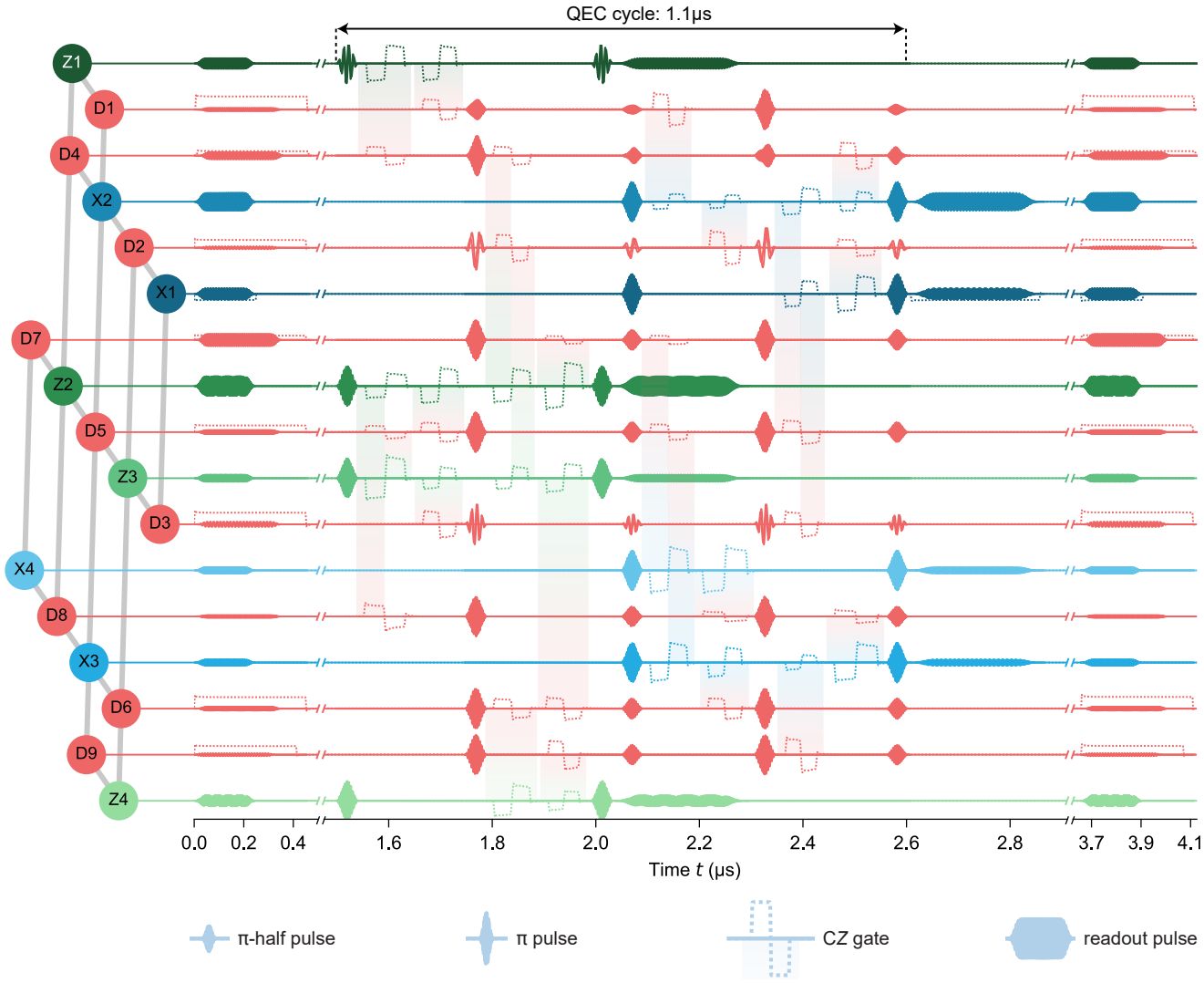

FIG. S12. AWG pulse sequence for the execution of a single quantum error correction cycle. Single-qubit drive pulses (solid lines), readout pulses (solid lines), and flux pulses (dotted lines) are displayed for each of the 17 qubits, as indicated by the label in the conceptual device representation. Qubit pairs realizing CZ gates are connected by shaded gradients. The time axis is segmented (parallel lines intersecting the axes) during idle times for clarity, see text for details.

data qubits to prepare the nine data qubits in one of the four product states $|0\rangle^{\otimes 9}$, $\hat{X}_{\mathrm{L}}|0\rangle^{\otimes 9}$, $|+\rangle^{\otimes 9}$ or $\hat{Z}_{\mathrm{L}}|+\rangle^{\otimes 9}$, each of which is a superposition of the 16 equivalent instances of the respective target logical state. Thereafter, Z-type stabilizers are executed by applying three simultaneous CZ gates in four sequential time steps, with a dynamical decoupling pulse applied to all data qubits in between the second and the third two-qubit gate time step. Each CZ gate is realized by applying a flux pulse to a data qubit and to the corresponding auxiliary qubit, see the shaded gradient connecting pairs of qubits in Fig. S12. While the Z-type auxiliary qubits are being read out, the X-type stabilizers are realized in a similar fashion as the Z-type stabilizers but with $\pi/2$-rotations implementing basis changes before and after the four time steps of CZ gates. Finally, all qubits are read out in the Z-basis after a latency of 800 ns set by the minimum re-triggering period of our acquisition devices.

We note that the duration of the error correction cycle can, in principle, be shortened using parallel scheduling of $\hat{S}^{\mathrm{X}i}$ and $\hat{S}^{\mathrm{Z}i}$ [14, 15] in case the readout operation is faster than the duration of the gate sequence of the stabilizers. In our experiment, the readout acquisition duration is 400 ns while the duration of the gate sequence is 550 ns. Hence, the error correction cycle duration could be reduced from 1.1 µs (pipelined scheduling [16] used in this work) to 0.95 µs when using parallel scheduling. However, we do not make use of parallel scheduling because it entails executing up to six CZ gates in parallel in a given time step, which imposes tighter constraints on the choice of two-qubit gate interaction frequencies.

Specifically, in our two-qubit gate scheme, we need

to make sure that the interaction frequencies of gates involving neighboring qubits are sufficiently distinct so that coherent errors induced by spectator-qubits remain small [35]. In addition, the interaction frequencies of a spatial sequence of parallel executed, neighboring gates need to be monotonously increasing (when starting the sequence with the auxiliary qubit of the first gate and ending it with the data qubit of the last gate) to avoid that qubits participating in neighboring gates cross frequencies. With parallel scheduling we would need four distinct interaction frequencies for a given time step (as opposed to only two distinct interaction frequencies for pipelined scheduling). In addition, a tight interaction frequency configuration leaves little range for individually adjusting interaction frequencies in the presence of defect modes on our device.

XI. LEAKAGE REJECTION

While non-computational states can provide a useful resource for example to realize two-qubit gates in transmon qubits, uncontrolled leakage into non-computational states constitutes a source of errors. Leakage has therefore been addressed both theoretically and experimentally by reducing leakage errors [4, 36], by detecting leakage indirectly on both auxiliary and data qubits using hidden Markov models [37] or quantum non-demolition measurement protocols [38], by developing leakage-aware decoding schemes [39, 40], and by converting leakage errors into errors within the computational subspace rendering them correctable in the standard error correction framework [41–43].

For our experimental realization of the surface code we mitigate leakage, particularly of data qubits, by choosing a frequency configuration and two-qubit gate scheme in which only the auxiliary qubits evolve through the second excited state $|2\rangle$, keeping the average leakage probability per data qubit and per cycle as low as about two per mill. Furthermore, we detect the remaining leakage events on both the data and auxiliary qubits by implementing a high-fidelity three-state readout scheme (Section VI), which allows us to reject all instances of experimental runs with detected leakage events and to study the performance of quantum error correction independent of leakage dynamics [44].

When analyzing the state preservation experiments shown in Fig. 4 of the main text, we find that the retained fraction of runs r in which no leakage event has been detected, decreases to good approximation exponentially with n, indicating that the retained fraction of runs per cycle r_c is independent of the cycle number, see Fig. S13. The value $r_c \approx 92.1(3)\,\%$, which we obtain from exponential fits $r = Ar_c^n$ (black lines) to the data (dark symbols) is identical within error bars for prepared eigenstates of $\hat{Z}_L$ (panel a) and $\hat{X}_L$ (panel b). We attribute the slight deviation between the measured data and the exponential fit to the finite probability of falsely

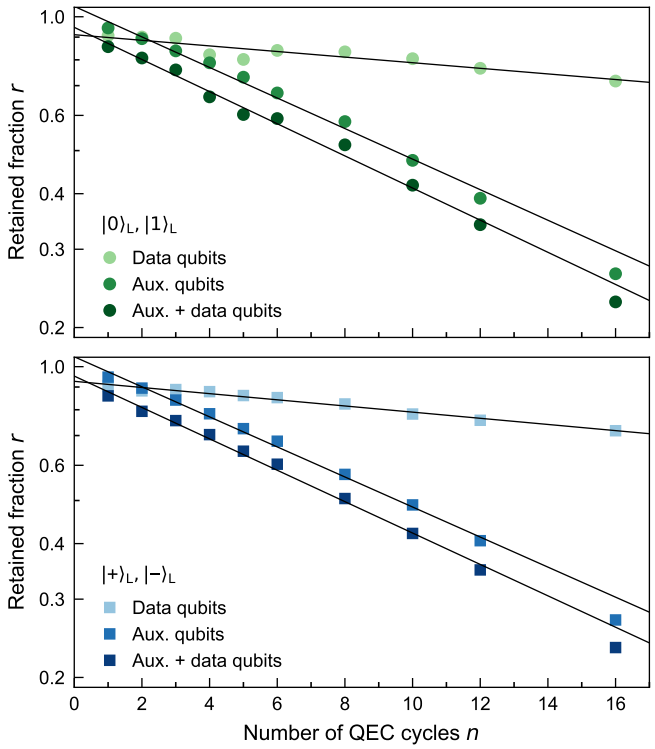

FIG. S13. Retained fraction of experimental runs after leakage rejection. **a** Retained fraction when rejecting leakage of data qubits only (light, green symbols), of auxiliary qubits only (mid-green), and of both (dark green). Solid black lines are exponential fits to the data. Data corresponds to the state preservation experiments of $|0\rangle_L$ and $|1\rangle_L$ presented in the main text. **b** Same as **a**, but for data from the state preservation experiments of $|+\rangle_L$ and $|-\rangle_L$.

classifying a $|1\rangle$-state as $|2\rangle$ during readout, which for auxiliary qubits is on average 0.004 and for data qubits 0.019. In fact, the probability for having such falsely identified leakage events is proportional to the population of the $|1\rangle$-state, which for auxiliary qubits increases from cycle to cycle due to their initialization in $|0\rangle$ and the error correction protocol, approaching 0.5 in the limit of large n. Hence, the falsely identified auxiliary-qubit leakage events increase with n, approaching about 21 % of the totally detected auxiliary qubit leakage events and explaining the slight increase of detected leakage events per cycle with n.

Distinguishing between leakage events on data and auxiliary qubits (lighter data points in Fig. S13 a, b), we find that the probability for having detected a leakage event on any of the nine data qubits is only about 0.015(2) per cycle and about 0.075(3) for the eight auxiliary qubits.

For comparison, we also determine the extracted logical error probabilities ϵ_L without leakage rejection (i), when rejecting data qubit leakage only (ii), and when rejecting auxiliary qubit leakage only (iii), see Table SIII. For simplicity, we interpret non-rejected measured $|2\rangle$-

TABLE SIII. Extracted logical error per cycle ϵ_{L} for the experiment preserving eigenstates of $\hat{Z}_{\mathrm{L}}$ and of $\hat{X}_{\mathrm{L}}$ using the indicated leakage rejection schemes. The retained data fraction per cycle r_c after leakage rejection is also indicated.

Leakage rejection	$\epsilon_{\mathrm{L}}\,[\hat{Z}_{\mathrm{L}}]$	$\epsilon_{\mathrm{L}}\,[\hat{X}_{\mathrm{L}}]$	r_c
(i) None	0.054(1)	0.049(2)	1.000(0)
(ii) Data qubits only	0.049(1)	0.044(1)	0.985(2)
(iii) Aux. qubits only	0.033(2)	0.030(1)	0.925(3)
(iv) Aux. and data qubits	0.032(1)	0.029(1)	0.921(3)

states as $|1\rangle$-states during decoding. Compared to the case when rejecting all detected leakage events (iv), the average absolute increase in the logical error probability is 0.021(2) for scheme (i), 0.016(1) for scheme (ii), and only 0.001(1) for scheme (iii). These results suggest that the low leakage rates achieved on our device, when combined with auxiliary-qubit reset using either feedback in combination with three-state readout or an unconditional scheme [22, 45], could render leakage errors tractable in general.

XII. DECODING AND WEIGHT EXTRACTION

To identify the most likely sequence of errors having occurred in a single instance of an experimental run, we decode the measured set of syndromes in post-processing by adopting the minimum-weight-perfect matching (MWPM) algorithm described in Ref. [30]. We represent each syndrome element σ_m^{Ai} as a vertex of a graph, which has two spatial dimensions set by the layout of auxiliary qubits Ai in the surface code lattice, and one temporal dimension indexed by the cycle number m. We connect pairs of vertices with edges, each of which represents a particular error at the physical level [40].

In our model, we consider three kinds of errors. First, auxiliary qubit errors, including errors during readout, which are represented by vertices at the same location but separated in time by one cycle ($\Delta m = 1$). Second, auxiliary qubit measurement misclassification errors, i.e. measurement errors which do not change the state of the auxiliary qubit, are represented by edges connecting vertices at the same location but separated in time by two cycles ($\Delta m = 2$). And third, single Pauli errors acting on data qubits, which are represented by edges connecting vertices either to a direct neighbor of the same type (X or Z) or to a boundary. Depending on the time at which these errors occur, the corresponding edge either connects to a vertex of the same cycle ($\Delta m = 0$) or the next cycle ($\Delta m = 1$). We extract the probabilities associated with those edges directly from measured syndrome correlations using the methods described in Refs. [46, 47]. Details of this scheme will be provided in a separate publication.

Based on the individual edge error probabilities, we then compute for *all* pairs of vertices k and l with $\Delta m \leq 2$, the total probability p_{kl} of being connected. We do so, by summing the individual probabilities over all possible error paths along the edges of the graph, see Ref. [46] for details. Here, the terms p_{kk} correspond to the probability of having vertex k being connected to a boundary. We convert the matrix of probabilities p_{kl} into a weight matrix $w_{kl} = -\ln p_{kl}$, based on which the MWPM algorithm connects each syndrome element $\sigma_m^{Ai} = 1$ either to a second such syndrome element or to a boundary while minimizing the total weight associated with these connections.

To correct for decoded errors in the final outcome z_{L} (x_{L}) of the logical operator $\hat{Z}_{\mathrm{L}}$ ($\hat{X}_{\mathrm{L}}$), we evaluate $z_{\mathrm{L}} = z_1 z_2 z_3$ ($x_{\mathrm{L}} = x_1 x_4 x_7$) from the final data qubit readout and multiply it by $(-1)^M$, where M is the number of syndrome element pairs determined by the MPWM algorithm, which signal a logical error. A syndrome element pair signals a logical error if the underlying error path contains an odd number of errors on those data qubits, which are contained in the logical operator string $\hat{Z}_{\mathrm{L}} = \hat{Z}_1\hat{Z}_2\hat{Z}_3$ ($\hat{X}_{\mathrm{L}} = \hat{X}_1\hat{X}_4\hat{X}_7$).

For consistency we evaluate the weights for each of the experimental and simulated state preservation experiments independently. We verified that the decoding with weights extracted from separate data sets yields the same logical lifetimes.

XIII. STATE PRESERVATION USING ERROR DETECTION

For comparison with the performance of our error-detection experiments in a distance-two surface-code [29], we estimate the logical life times in an error detection setting, i.e. when postselecting on experimental runs in which all syndrome elements are zero. We restrict the analysis to a maximum of six quantum error correction cycles to retain at least 500 experimental runs per data point after postselection.

Given the limited amount of data available, we do not resolve any discernible decay of $\langle\hat{Z}_{\mathrm{L}}\rangle$ and $\langle\hat{X}_{\mathrm{L}}\rangle$ when postselecting on no detected errors, see semi-filled symbols in Fig. S14 a, b. The decay is clearly much less than the one observed when correcting errors, see filled data points in Fig. S14 a, b (same data as in Fig. 4 a, b). With the statistical scatter of the data points being smaller than the expected decay at a logical life time of 1 ms, we roughly estimate logical lifetimes of at least 1 ms of our logical qubit in the error detection setting.

Accordingly, we find a much reduced logical error probability E_{L} (semi-filled symbols in Fig. S14 c). As a reference, we also display the bare logical operator expectation values – evaluated executing the error correction cycles but not making use of the measurement results neither for error detection nor correction – and the corresponding logical error probability on the same scales (open symbols in Fig. S14). The logical life times extracted for the bare, the error-corrected and the error-

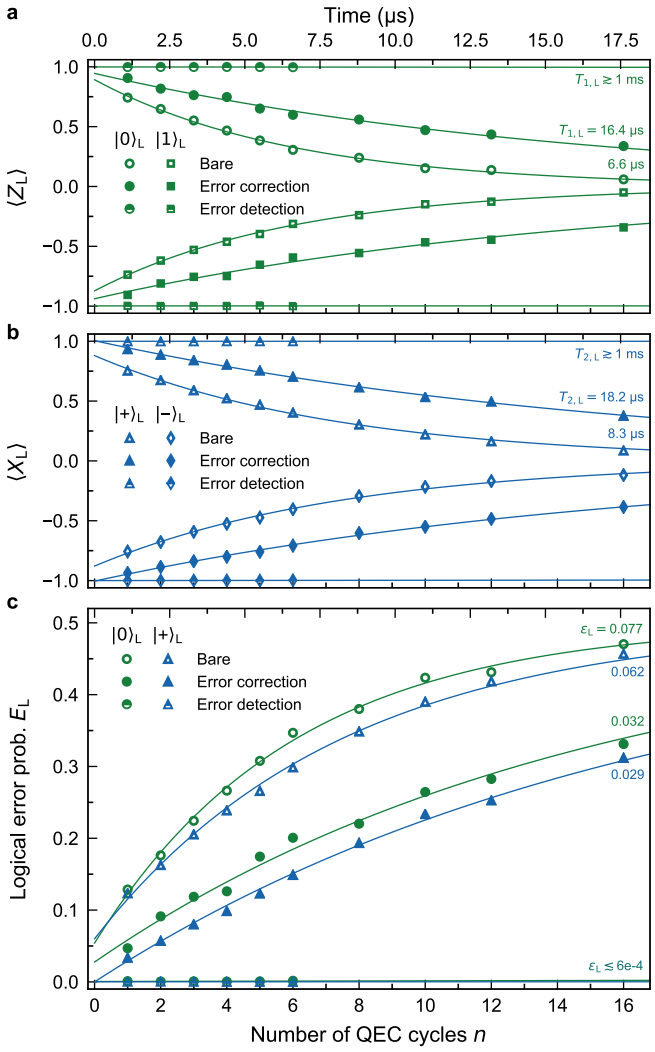

FIG. S14. Comparison of the performance of logical state preservation when employing no error correction (bare, open symbols), error correction (filled symbols) and error detection (semi-filled symbols), see text for details. **a** Expectation value of the $\hat{Z}_L$ operator for prepared $|0\rangle_L$ (circles) and $|1\rangle_L$ (squares) as a function of the number n of error correction cycles. Exponential fits are shown as solid lines. The extracted decay times are indicated on the right. **b** Corresponding data sets for $\hat{X}_L$, $|+\rangle_L$ (triangles) and $|-\rangle_L$ (diamonds). **c** Logical error probability E_L for $|0\rangle_L$ and $|+\rangle_L$ and the extracted error per cycle ϵ_L indicated on the right.

detected data are summarized in Tab. SIV.

The logical coherence times when using error detection largely surpass any of the physical qubit coherence times and thus demonstrate the effectiveness of detecting errors using high-fidelity auxiliary-qubit-based stabilizer measurements. The main reason for the largely increased logical life times compared to the error correction setting is that logical errors remain undetected only if error events resulting in three or more errors on data qubits change the logical state without generating any non-trivial syndrome element $\sigma_{Ai} = 1$.

As expected, the number of experimental runs kept after postselection in the error detection setting decreases exponentially as a function of n, reaching ~ 700 runs per prepared logical state at $n = 6$. Prior to the error detection procedure which postselects on runs which have only syndrome elements of zero, the available data has already been reduced by ground state heralding and leakage rejection. The error detection procedure itself discards about 50 % of the (already reduced) data per cycle.

TABLE SIV. Extracted logical life times $T_{1,L}$ and logical coherence times $T_{2,L}$ for the experiments preserving eigenstates of $\hat{Z}_L$ and of $\hat{X}_L$, respectively, for the indicated data treatments.

Data treatment	$T_{1,L}$	$T_{2,L}$
Bare	6.6(2) μs	8.3(2) μs
Error correction	16.4(8) μs	18.2(5) μs
Error detection	$\gtrsim 1$ ms	$\gtrsim 1$ ms

[1] A. Dunsworth, A. Megrant, C. Quintana, Z. Chen, R. Barends, B. Burkett, B. Foxen, Y. Chen, B. Chiaro, A. Fowler, R. Graff, E. Jeffrey, J. Kelly, E. Lucero, J. Y. Mutus, M. Neeley, C. Neill, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, and J. M. Martinis, Characterization and reduction of capacitive loss induced by sub-micron josephson junction fabrication in superconducting qubits, Appl. Phys. Lett. **111**, 022601 (2017).

[2] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, Phys. Rev. Lett. **103**, 110501 (2009).

[3] M. A. Rol, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, and L. DiCarlo, Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits,

Phys. Rev. Lett. **123**, 120502 (2019).
[4] V. Negirneac, H. Ali, N. Muthusubramanian, F. Battistel, R. Sagastizabal, M. S. Moreira, J. F. Marques, W. J. Vlothuizen, M. Beekman, C. Zachariadis, N. Haider, A. Bruno, and L. DiCarlo, High-fidelity controlled-z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor, Phys. Rev. Lett. **126**, 220502 (2021).
[5] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).
[6] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, Investigating the limits of randomized benchmarking protocols, Phys. Rev. A **89**, 062321 (2014).
[7] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).
[8] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, and M. Steffen, Process verification of two-qubit quantum gates by randomized benchmarking, arXiv:1210.7011 (2012).
[9] 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. OMalley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature **508**, 500 (2014).
[10] F. W. Strauch, P. R. Johnson, A. J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood, Quantum logic gates for coupled superconducting phase qubits, Phys. Rev. Lett. **91**, 167005 (2003).
[11] L. DiCarlo, M. D. Reed, L. Sun, B. R. Johnson, J. M. Chow, J. M. Gambetta, L. Frunzio, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf, Preparation and measurement of three-qubit entanglement in a superconducting circuit, Nature **467**, 574 (2010).
[12] M. A. Rol, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, R. E. Sagastizabal, C. C. Bultink, Y. Salathe, N. Haandbaek, J. Sedivy, and L. DiCarlo, Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor, Appl. Phys. Lett. **116**, 054001 (2020).
[13] C. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, Surface code quantum computing by lattice surgery, New Journal of Physics **14**, 123011 (2012).
[14] Y. Tomita and K. M. Svore, Low-distance surface codes under realistic quantum noise, Phys. Rev. A **90**, 062320 (2014).
[15] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).
[16] R. Versluis, S. Poletto, N. Khammassi, B. Tarasinski, N. Haider, D. J. Michalak, A. Bruno, K. Bertels, and L. DiCarlo, Scalable quantum circuit and control for a superconducting surface code, Phys. Rev. Applied **8**, 034021 (2017).
[17] S. Krinner, S. Storz, P. Kurpiers, P. Magnard, J. Heinsoo, R. Keller, J. Lütolf, C. Eichler, and A. Wallraff, Engineering cryogenic setups for 100-qubit scale superconducting circuit systems, EPJ Quantum Technology **6**, 2 (2019).
[18] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi, A near-quantum-limited Josephson traveling-wave parametric amplifier, Science **350**, 307 (2015).
[19] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, and C. Eichler, Rapid high-fidelity multiplexed readout of superconducting qubits, Phys. Rev. Appl. **10**, 034040 (2018).
[20] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Rapid, high-fidelity, single-shot dispersive readout of superconducting qubits, Phys. Rev. Appl. **7**, 054020 (2017).
[21] P. Kurpiers, P. Magnard, T. Walter, B. Royer, M. Pechal, J. Heinsoo, Y. Salathé, A. Akin, S. Storz, J.-C. Besse, S. Gasparinetti, A. Blais, and A. Wallraff, Deterministic quantum state transfer and remote entanglement using microwave photons, Nature **558**, 264 (2018).
[22] P. Magnard, P. Kurpiers, B. Royer, T. Walter, J.-C. Besse, S. Gasparinetti, M. Pechal, J. Heinsoo, S. Storz, A. Blais, and A. Wallraff, Fast and unconditional all-microwave reset of a superconducting qubit, Phys. Rev. Lett. **121**, 060502 (2018).
[23] N. Lacroix, C. Hellings, C. K. Andersen, A. Di Paolo, A. Remm, S. Lazar, S. Krinner, G. J. Norris, M. Gabureac, J. Heinsoo, A. Blais, C. Eichler, and A. Wallraff, Improving the performance of deep quantum optimization algorithms with continuous gate sets, PRX Quantum **1**, 110304 (2020).
[24] J. Dalibard, Y. Castin, and K. Mølmer, Wave-function approach to dissipative processes in quantum optics, Physical review letters **68**, 580 (1992).
[25] K. Mølmer, Y. Castin, and J. Dalibard, Monte carlo wave-function method in quantum optics, JOSA B **10**, 524 (1993).
[26] R. Dum, P. Zoller, and H. Ritsch, Monte Carlo simulation of the atomic master equation for spontaneous emission, Physical Review A **45**, 4879 (1992).
[27] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP: An open-source Python framework for the dynamics of open quantum systems, Computer Physics Communications **183**, 1760 (2012).
[28] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient z gates for quantum computing, Phys. Rev. A **96**, 022330 (2017).
[29] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, Repeated quantum error detection in a surface code, Nature Physics **16**, 875 (2020).
[30] T. E. O'Brien, B. Tarasinski, and L. DiCarlo, Density-matrix simulation of small surface codes under current and projected experimental noise, npj Quantum Inf. **3**, 39 (2017).
[31] C. Huang, X. Ni, F. Zhang, M. Newman, D. Ding, X. Gao, T. Wang, H. Zhao, F. Wu, G. Zhang, C. Deng, H. Ku, J. Chen, and Y. Shi, Alibaba cloud quantum development platform: Surface code simulations with crosstalk, arXiv:2002.08918 (2020).
[32] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
[33] D. Nigg, M. Müller, E. A. Martinez, P. Schindler,

M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt, Quantum computations on a topologically encoded qubit, Science **345**, 302 (2014).

[34] M. H. Abobeih, Y. Wang, J. Randall, S. J. H. Loenen, C. E. Bradley, M. Markham, D. J. Twitchen, B. M. Terhal, and T. H. Taminiau, Fault-tolerant operation of a logical qubit in a diamond quantum processor, arXiv:2108.01646 (2021).

[35] S. Krinner, S. Lazar, A. Remm, C. Andersen, N. Lacroix, G. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, Benchmarking coherent errors in controlled-phase gates due to spectator qubits, Phys. Rev. Appl. **14**, 024042 (2020).

[36] Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, and J. M. Martinis, Measuring and suppressing quantum state leakage in a superconducting qubit, Phys. Rev. Lett. **116**, 020501 (2016).

[37] C. C. Bultink, T. E. O'Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, M. A. Rol, X. Fu, B. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, and L. DiCarlo, Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements, Science Advances **6**, eaay3050 (2020).

[38] R. Stricker, D. Vodola, A. Erhard, L. Postler, M. Meth, M. Ringbauer, P. Schindler, T. Monz, M. Müller, and R. Blatt, Experimental deterministic correction of qubit loss, Nature **585**, 207 (2020).

[39] M. Suchara, A. W. Cross, and J. M. Gambetta, Quantum Info. Comput. **15**, 997 (2015).

[40] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, State preservation by repetitive error detection in a superconducting quantum circuit, Nature **519**, 66 (2015).

[41] P. Aliferis and B. M. Terhal, Fault-tolerant quantum computation for local leakage faults, Quantum Info. Comput. **7**, 139 (2007).

[42] A. G. Fowler, Coping with qubit leakage in topological codes, Phys. Rev. A **88**, 042308 (2013).

[43] J. Ghosh and A. G. Fowler, Leakage-resilient approach to fault-tolerant quantum computing with superconducting elements, Physical Review A **91**, 020302 (2015).

[44] B. M. Varbanov, F. Battistel, B. M. Tarasinski, V. P. Ostroukh, T. E. O'Brien, L. DiCarlo, and B. M. Terhal, Leakage detection for a transmon-based surface code, npj Quantum Information **6**, 102 (2020).

[45] M. McEwen, D. Kafri, Z. Chen, J. Atalaya, K. J. Satzinger, C. Quintana, P. V. Klimov, D. Sank, C. Gidney, A. G. Fowler, F. Arute, K. Arya, B. Buckley, B. Burkett, N. Bushnell, B. Chiaro, R. Collins, S. Demura, A. Dunsworth, C. Erickson, B. Foxen, M. Giustina, T. Huang, S. Hong, E. Jeffrey, S. Kim, K. Kechedzhi, F. Kostritsa, P. Laptev, A. Megrant, X. Mi, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Niu, A. Paler, N. Redd, P. Roushan, T. C. White, J. Yao, P. Yeh, A. Zalcman, Y. Chen, V. N. Smelyanskiy, J. M. Martinis, H. Neven, J. Kelly, A. N. Korotkov, A. G. Petukhov, and R. Barends, Removing leakage-induced correlated errors in superconducting quantum error correction, Nature Communications **12**, 1 (2021).

[46] S. T. Spitz, B. Tarasinski, C. W. J. Beenakker, and T. E. O'Brien, Adaptive weight estimator for quantum error correction in a time-dependent environment, Advanced Quantum Technologies **1**, 1800012 (2018).

[47] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, S. Hong, C. Jones, A. Petukhov, D. Kafri, S. Demura, B. Burkett, C. Gidney, A. G. Fowler, A. Paler, H. Putterman, I. Aleiner, F. Arute, K. Arya, R. Babbush, J. C. Bardin, A. Bengtsson, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, N. Bushnell, B. Chiaro, R. Collins, W. Courtney, A. R. Derk, D. Eppens, C. Erickson, E. Farhi, B. Foxen, M. Giustina, A. Greene, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, T. Huang, W. J. Huggins, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, K. Kechedzhi, S. Kim, A. Kitaev, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, T. McCourt, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Newman, M. Y. Niu, T. E. O'Brien, A. Opremcak, E. Ostby, B. Pató, N. Redd, P. Roushan, N. C. Rubin, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. White, Z. J. Yao, P. Yeh, J. Yoo, A. Zalcman, H. Neven, S. Boixo, V. Smelyanskiy, Y. Chen, A. Megrant, J. Kelly, and A. I. Google Quantum, Exponential suppression of bit or phase errors with cyclic error correction, Nature **595**, 383 (2021).