ARTICLE   OPEN

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

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

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

INTRODUCTION

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

In this paper, we perform a density-matrix simulation of the distance-3 surface code named Surface-17, using the concrete quantum circuit recently proposed in ref. 18 and the measured performance of current experimental multi-transmon cQED platforms.[19–22] For this purpose, we have developed an open-source density-matrix simulation package named quantumsim (https://github.com/brianzi/quantumsim). We use quantumsim to extract the logical error rate per QEC cycle, ϵ_L. This metric allows us to

optimize and trade-off between QEC cycle parameters, assess the merits of feedback control, predict gains from future improvements in physical qubit performance, and quantify decoder performance. We compare an algorithmic decoder using minimum-weight perfect matching (MWPM) with homemade weight calculation to a simple look-up table (LT) decoder, and weigh both against an upper bound (UB) for decoder performance obtainable from the density-matrix simulation. Finally, we make a low-order approximation to extend our predictions to the distance-5 Surface-49. The combination of results for Surface-17 and Surface-49 allows us to make statements about code scaling and to predict the code size and physical qubit performance required to achieve break-even points for memory and computational performance.

RESULTS

Error rates for Surface-17 under current experimental conditions

To quantify the performance of the logical qubit, we first define a test experiment to simulate. Inspired by the recent experimental demonstration of distance-3 and distance-5 repetition codes,[4] we first focus on the performance of the logical qubit as a quantum memory. Specifically, we quantify the ability to hold a logical $|0\rangle$ state, by initializing this state, holding it for $k \in \{1, ..., 20\}$ cycles, performing error correction, and determining a final logical state (see Fig. 6 for details). The logical fidelity $\mathcal{F}_L[k]$ is then given by the probability to match the initial state. We observe identical results when using $|1\rangle$ or $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ in place of $|0\rangle$.

We base our error model for the physical qubits on current typical experimental performance for transmons in planar cQED, using parameters from the literature and in-house results (e.g., gate-set tomography measurements). These are summarized in Table 1, and further detailed in Supplemental Material. We focus

[1]Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands; [2]QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands and [3]Kavli Institute of Nanoscience, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands
Correspondence: L. DiCarlo (l.dicarlo@tudelft.nl)
T. E. O'Brien and B. Tarasinski contributed equally to this work.

npj

2

Density-matrix simulation of small surface
TE O'Brien et al.

Table 1. Standard simulation parameters

Parameter	Symbol	Value	Reference
Qubit relaxation time	T_1	30 μs	19
Qubit dephasing time (white noise)	T_ϕ	60 μs	19, 21
Single-qubit gate time	$T_{g,1Q}$	20 ns	19, 21
Two-qubit gate time	$T_{g,2Q}$	40 ns	5
Coherent step time	τ_c	200 ns	18
Measurement time	τ_m	300 ns	19
Depletion time	τ_d	300 ns	19
Fast measurement time	$\tau_m^{(fast)}$	100 ns	22
Fast depletion time	$\tau_d^{(fast)}$	100 ns	22

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

on the QEC cycle proposed in ref. 18 which pipelines the execution of X-type and Z-type stabilizer measurements. Each stabilizer measurement consists of three parts: a coherent step (duration $\tau_c = 2\tau_{g,1Q} + 4\tau_{g,2Q}$), measurement ($\tau_m$), and photon depletion from readout resonators (τ_d), making the QEC cycle time $\tau_{cycle} = \tau_c + \tau_m + \tau_d$.

Simulating this concrete quantum circuit with the listed parameters using quantumsim, we predict $\mathcal{F}_L[k]$ of Surface-17 (Fig. 1). We show $\mathcal{F}_L[k]$ for both a homemade MWPM decoder (green, described in Supplemental Material), and an implementation of the LT decoder of ref. 13 (blue, described in Supplemental Material). To isolate decoder performance, we can compare the achieved fidelity to an UB extractable from the density-matrix simulation (red, described in Methods). To assess the benefit of QEC, we also compare to a single decohering transmon, whose fidelity is calculated by averaging over the six cardinal points of the Bloch sphere:

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

The observation of $\mathcal{F}_L[k] > \mathcal{F}_{phys}(k\tau_{cycle})$ for large k would constitute a demonstration of QEC beyond the quantum memory break-even point.[7] Equivalently, one can extract a logical error rate ϵ_L from a best fit to $\mathcal{F}_L[k]$ (as derived in Methods as the probability of an odd number of errors occurring),

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

Here, k_0 and ϵ_L are the parameters to be fit. We compare ϵ_L to the physical error rate

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

We observe $\epsilon_L = 1.44\%_c$ for the LT decoder, $\epsilon_L = 1.07\%_c$ for the MWPM decoder, and $\epsilon_L = 0.68\%_c$ at the decoder UB ($\%_c = \%$ per cycle). The latter two fall below $\epsilon_{phys} = 1.33\%_c$. Defining the decoder efficiency $\eta_d = \epsilon_L^{(UB)}/\epsilon_L$, we find $\eta_d^{(LT)} = 0.47$, and $\eta_d^{(MWPM)} = 0.64$.

We can also compare the multi-cycle error correction to majority voting, in which the state declaration is based solely on the output of the final data qubit measurements (ancilla measurements are ignored). Majority voting corrects any single data qubit error (over the entire experiment), and thus exhibits a quadratic decay for small k (A distance-d code with majority voting alone should exhibit a $(d+1)/2$-order decay). A decoder should also be able to correct (at least) a single error, and thus should produce the same behavior at low k, delaying the onset of exponential decay in $\mathcal{F}_L[k]$. In fact, a good test for the performance of a MWPM decoder is to ensure it can outperform
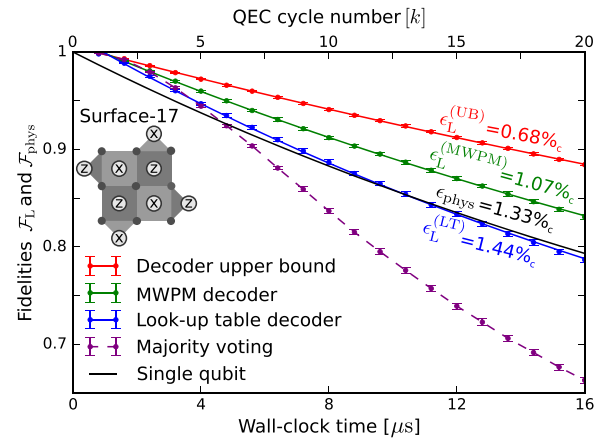

Fig. 1 Logical fidelity $\mathcal{F}_L[k]$ of Surface-17 with current experimental parameters (Table 1 and (See Supplemental Material)), simulated with quantumsim as described in Fig. 6. The results from a MWPM decoder (green) and an implementation of the LT decoder of ref. 13 (blue) are compared to the decoder upper bound (red). The labeled error rate is obtained from the best fit to Eq. (2) (also plotted). A further comparison is given to majority voting (purple, dashed), which ignores the outcome of individual stabilizer measurements, and to the fidelity $\mathcal{F}_{phys}$ of a single transmon (black) [Eq. (1)]. Error bars (2 s.d.) are obtained by bootstrapping

the majority vote at short timescales, as suboptimal configuration will prevent this (as seen for the LT decoder).

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

Optimization of logical error rates with current experimental conditions

Error sources in current cQED setups derive primarily from transmon decoherence, as opposed to gate and measurement errors produced by control electronics. Thus, a path to reducing ϵ_L may be to decrease τ_{cycle}. Currently, the cycle is dominated by $\tau_m + \tau_d$. At fixed readout power, reducing τ_m and τ_d will reduce τ_{cycle} at the cost of increased readout infidelity ϵ_{RO} (described in Methods). We explore this trade-off in Fig. 2, using a linear-dispersive readout model,[23] keeping $\tau_m = \tau_d$ and assuming no leftover photons. Because of the latter, $\epsilon_L^{(MWPM)}$ reduces from 1.07%_c (Fig. 1) to 0.62%_c at $\tau_m = 300$ ns. The minimum $\epsilon_L^{(MWPM)} = 0.55\%_c$ is achieved at around $\tau_m = 260$ ns. This is perhaps

Density-matrix simulation of small surface
TE O'Brien et al.

npj

3
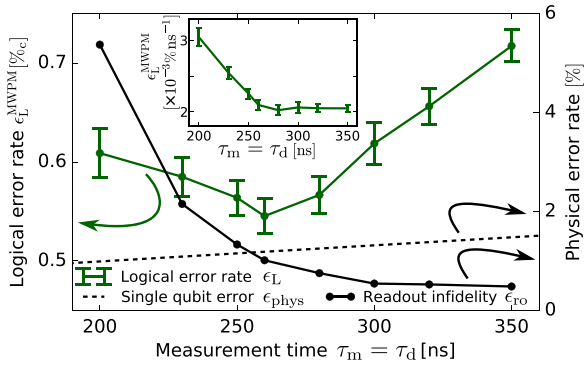
Fig. 2 Optimization of the logical error rate (per cycle) of Surface-17 as a function of measurement-and-depletion time.[19] Changes in the underlying physical error rates are shown, as well. Decreasing the measurement time causes an increase in the readout infidelity (solid black curve with dots), whilst decreasing the single qubit decay from T_1 and T_2 (black dashed curve) for all qubits. The logical rate with an MWPM decoder (green curve) is minimized when these error rates are appropriately balanced. The logical error rate is calculated from the best fit of Eq. (2). Error bars (2 s.d.) are obtained by bootstrapping ($N = 10,000$ runs). Inset: Logical error rate per unit time, instead of per cycle

counterintuitive, as ϵ_{phys} reduces only 0.13%$_c$ while ϵ_{RO} increases 0.5%. However, it reflects the different sensitivity of the code to different types of errors. Indeed, $\epsilon_L^{(MWPM)}$ is smaller for $\tau_m = 200$ ns than for $\tau_m = 300$ ns, even though ϵ_{RO} increases to 5%. It is interesting to note that the optimal τ_m for quantum memory, which minimizes logical error per unit time, rather than per cycle, is $\tau_m = 280$ ns (Fig. 2 inset). This shows that different cycle parameters might be optimal for computation and memory applications.

Next, we consider the possibility to reduce ϵ_L using feedback control. Since T_1 only affects qubits in the excited state, the error rate of ancillas in Surface-17 is roughly two times higher when in the excited state. The unmodified syndrome extraction circuit flips the ancilla if the corresponding stabilizer value is −1, and since ancillas are not reset between cycles, they will spend significant amounts of time in the excited state. Thus, we consider using feedback to hold each ancilla in the ground state as much as possible. We do not consider feedback on data qubits, as the highly entangled logical states are equally susceptible to T_1.

The feedback scheme (Inset of Fig. 3) consists of replacing the $R_y(\pi/2)$ gate at the end of the coherent step with a $R_y(-\pi/2)$ gate for some of the ancillas, depending on a classical control bit p for each ancilla. This bit p represents an estimate of the stabilizer value, and the ancilla is held in the ground state whenever this estimate is correct (i.e., in the absence of errors). Figure 3 shows the effect of this feedback on the logical fidelity, both for the MWPM decoder and the decoder UB. We observe ϵ_L improve only 0.05%$_c$ in both cases. Future experiments might opt not to pursue these small gains in view of the technical challenges added by feedback control.

Projected improvement with advances in quantum hardware

We now estimate the performance increase that may result from improving the transmon relaxation and dephasing times via materials and filtering improvements. To model this, we return to $\tau_{cycle} = 800$ ns, and adjust T_1 values with both $T_\phi = 2T_1$ (common in experiment) and $T_\phi = \infty$ (all white-noise dephasing eliminated). We retain the same rates for coherent errors, readout infidelity, and photon-induced dephasing as in Fig. 1. Figure 4 shows the extracted ϵ_L and ϵ_{phys} over the T_1 range covered. For the MWPM decoder (UB) and $T_\phi = 2T_1$, the memory figure of merit $\gamma_m = \epsilon_{phys}/\epsilon_L$ increases from 1.3 (2) at $T_1 = 30\,\mu s$ to 2 (5) at $100\,\mu s$.
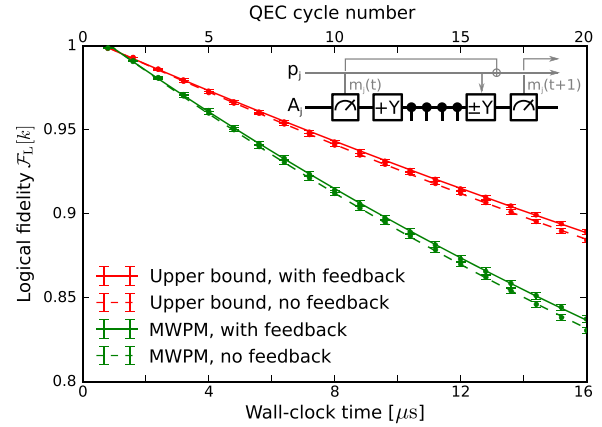
Fig. 3 Logical fidelity of Surface-17 with (solid) and without (dashed) an additional feedback scheme. The performance of a MWPM decoder (green) is compared to the decoder upper bound (red). Curves are fits of Eq. (2) to the data, and error bars (2 s.d.) are given by bootstrapping, with each point averaged over 10,000 runs. Inset: Method for implementing the feedback scheme. For each ancilla qubit A_j, we store a parity bit p_j, which decides the sign of the $R_y(\pi/2)$ rotation at the end of each coherent step. The time A_j spends in the ground state is maximized when p_j is updated each cycle t by XORing with the measurement result from cycle $t − 1$, after the rotation of cycle t has been performed
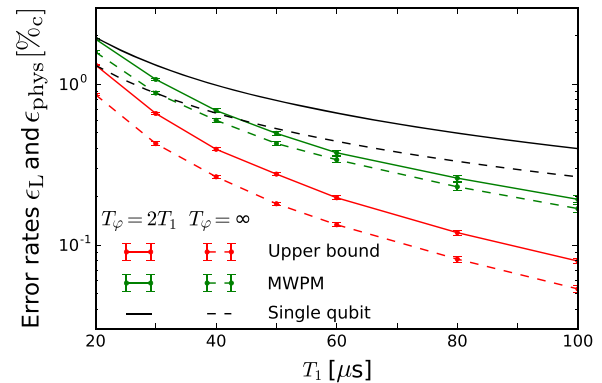

Fig. 4 T_1 dependence of the Surface-17 logical error rate (MWPM and UB) and the physical error rate. We either fix $T_\phi = 2T_1$ (solid) or $T_\phi = \infty$ (dashed). Logical error rates are extracted from a best fit of Eq. (2) to $\mathcal{F}_L[k]$ over $k = 1, …, 20$ QEC cycles, averaged over $N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

Completely eliminating white-noise dephasing will increase γ_m by 10% with MWPM and 30% at the UB.

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

As a rough metric of computational performance, we offer to compare ϵ_L (per cycle) to the error accrued by a physical qubit

npj

Density-matrix simulation of small surface
TE O'Brien et al.

4
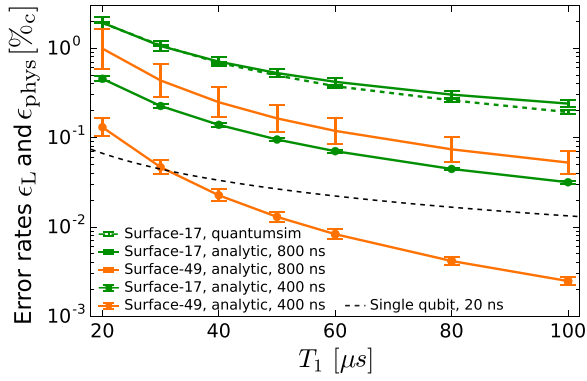
Fig. 5 Analytic approximation of ϵ_L for Surface-17 (green) and Surface-49 (orange) using a MWPM decoder. Details of the calculation of points and error bars are given in (See Supplemental Material). All plots assume $T_\phi = 2T_1$, and $\tau_{cycle} = 800$ ns (crosses) or 400 ns (dots). Numerical results for Surface-17 with $\tau_{cycle} = 800$ ns are also plotted for comparison (green, dashed). The physical-qubit computation metric is given as the error incurred by a single qubit over the resting time of a single-qubit gate (black, dashed)

idling over $\tau_{g,1Q}$. We define a metric for computation performance, $\gamma_c = (\epsilon_{phys}\tau_{g,1Q})/(\epsilon_L\tau_{cycle})$ and $\gamma_c = 1$ as a computational break-even point. Clearly, using the QEC cycle parameters of Table 1 and even with T_1 improvements, neither Surface-17 nor Surface-49 can break-even computationally. However, including the readout acceleration recently demonstrated in ref. 22 which allows $\tau_m = \tau_d = 100$ ns and $\tau_{cycle} = 400$ ns, Surface-49 can cross $\gamma_c = 1$ by $T_1 = 40$ μs . In view of first reports of T_1 up to 80 μs emerging for planar transmons,[24, 25] this important milestone may be within grasp.

DISCUSSION

Computational figure of merit

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

Decoder performance

A practical question facing QEC is how best to balance the trade-off between decoder complexity and performance. Past proposals for surface-code computation via lattice surgery[17] require the decoder to provide an up-to-date estimate of the Pauli error on physical qubits during each logical T gate. Because tracking Pauli errors through a non-Clifford gate is inefficient, however implemented, equivalent requirements will hold for any QEC code.[26] A decoder is thus required to process ancilla measurements from one cycle within the next (on average). This presents a considerable challenge for transmon-cQED implementations, as $\tau_{cycle} < 1$ μs. This short time makes the use of computationally intensive decoding schemes difficult, even if they provide lower ϵ_L.

The leading strategy for decoding the surface code is MWPM using the blossom algorithm of Edmonds.[10, 14, 27] Although this algorithm is challenging to implement, it scales linearly in code distance.[27] The algorithm requires a set of weights (representing the probability that two given error signals are connected by a chain of errors) as input. An important practical question

(See Supplemental Material) is whether these weights can be calculated on the fly, or must be precalculated and stored. On-the-fly weight calculation is more flexible. For example, it can take into account the difference in error rates between an ancilla measured in the ground and in the excited state. The main weakness of MWPM is the inability to explicitly detect Y errors. In fact, in Supplemental Material we show that MWPM is nearly perfect in the absence of Y errors. The decoder efficiency Y may significantly increase by extending MWPM to account for correlations between detected X and Z errors originating from Y errors.[28, 29]

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

Other observations

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

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

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

A final, interesting point for future surface-code computation is shown in Fig. 2: the optimal cycle parameters for logical error rates per cycle and per unit time are not the same. This implies that logical qubits functioning as a quantum memory should be treated differently to those being used for computation. This idea can be extended further: at any point in time, a large quantum computer performing a computation will have a set S_m of memory qubits which are storing part of a large entangled state, whilst a set S_c of computation qubits containing the rest of the state undergo operations. To minimize the probability of a logical error occurring on qubits within both S_c and S_m, the cycle time of the qubits in S_c can be reduced to minimize the rest time of qubits in S_m. As a simple example, consider a single computational qubit q_c and a single memory qubit q_m sharing entanglement. Operating all qubits at $\tau_{cycle} = 720$ ns to minimize ϵ_L would lead to a 1.09% error rate for the two qubits combined. However, shortening the τ_{cycle} of q_c reduces the time over which q_m decays. If q_c operates at

Density-matrix simulation of small surface
TE O'Brien et al.

npj

5

$\tau_{cycle} = 600$ ns, the average error per computational cycle drops to 1.06%, as q_m completes only 5 cycles for every 6 on q_c. Although this is only a meager improvement, one can imagine that when many more qubits are resting than performing computation, the relative gain will be quite significant.

Effects not taken into account

Although we have attempted to be thorough in the detailing of the circuit, we have neglected certain effects. We have used a simple model for C-Z gate errors as we lack data from experimental tomography (e.g., one obtained from two-qubit gate-set tomography[30]). Most importantly, we have neglected leakage, where a transmon is excited out of the two lowest energy states, i.e., out of the computational subspace. Previous experiments have reduced the leakage probability per C-Z gate to ~0.3%,[31] and per single-qubit gate to ~0.001%.[32] Schemes have also been developed to reduce the accumulation of leakage.[33] Extending quantumsim to include and investigate leakage is a next target. However, the representation of the additional quantum state can increase the simulation effort significantly [by a factor of $(9/4)^{10} \approx 3000$]. To still achieve this goal, some further approximations or modifications to the simulation will be necessary. Future simulations will also investigate the effect of

spread in qubit parameters, both in space (i.e., variation of physical error rates between qubits) and time (e.g., T_1 fluctuations), and cross-talk effects such as residual couplings between nearest and next-nearest neighbor transmons, qubit cross-driving, and qubit dephasing by measurement pulses targeting other qubits.

METHODS

Simulated experimental procedure

Surface-17 basics. A QEC code can be defined by listing the data qubits and the stabilizer measurements that are repeatedly performed upon them.[34] In this way, Surface-17 is defined by a 3×3 grid of data qubits {D_0, ... D_8}. In order to stabilize a single logical qubit, $9 - 1 = 8$ commuting measurements are performed. The stabilizers are the weight-two and weight-four X-type and Z-type parity operators $X_2 X_1$, $Z_3 Z_0$, $X_4 X_3 X_1 X_0$, $Z_5 Z_4 Z_2 Z_1$, $Z_7 Z_6 Z_4 Z_3$, $X_8 X_7 X_5 X_4$, $Z_8 Z_5$, and $X_7 X_6$, where X_j (Z_j) denotes the X (Z) Pauli operator acting on data qubit D_j. Their measurement is realized indirectly using nearest-neighbor interactions between data and ancilla qubits arranged in a square lattice, followed by ancilla measurements [Fig. 6a]. This leads to a total of 17 physical qubits when a separate ancilla is used for each individual measurement. We follow the circuit realization of this code described in ref. 18, for which we give a schematic description in Fig. 6b (See Supplemental Material for a full circuit diagram).

In an experimental realization of this circuit, qubits will regularly accumulate errors. Multiple errors that occur within a short period of time
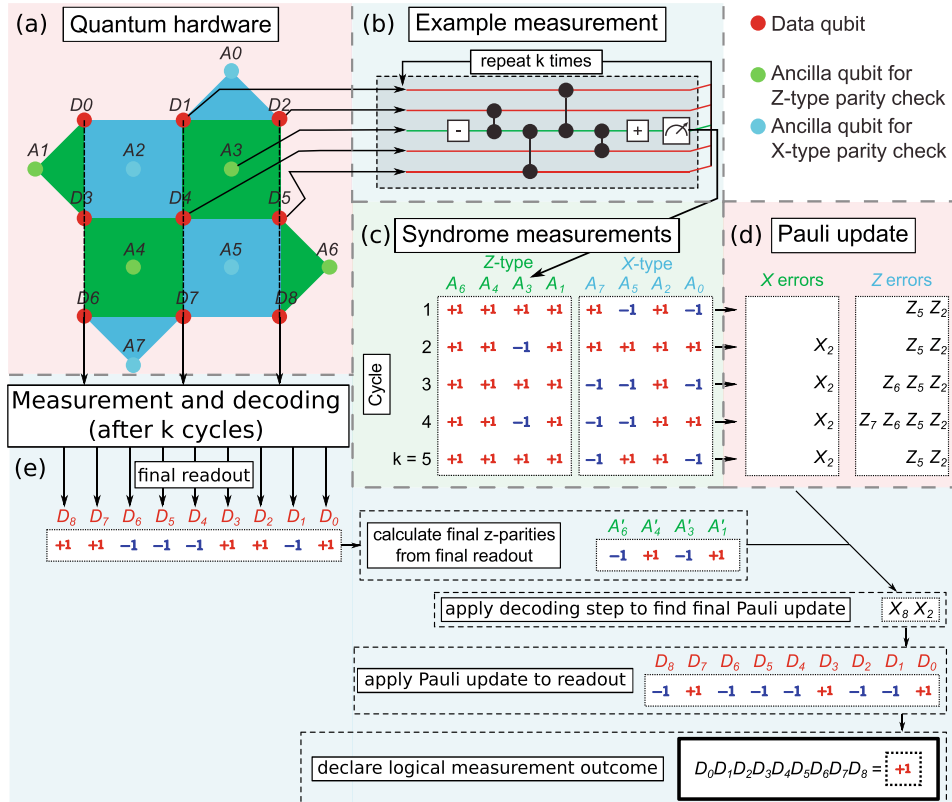
Fig. 6 Schematic overview of the simulated experiment. **a** 17 qubits are arranged in a surface code layout (legend top-right). The red data qubits are initialized in the ground state $|0\rangle$, and projected into an eigenstate of the measured X-(blue) and Z-(green) type stabilizer operators. **b** A section of the quantum circuit depicting the four-bit parity measurement implemented by the A_3 ancilla qubit (+/− refer to $R_y(\pm\pi/2)$ single-qubit rotations). The ancilla qubit (green line, middle) is entangled with the four data qubits (red lines) to measure $Z_1 Z_2 Z_4 Z_5$. Ancillas are not reset between cycles. Instead, the implementation relies on the quantum non-demolition nature of measurements. The stabilizer is then the product of the ancilla measurement results of successive cycles. This circuit is performed for all ancillas and repeated k times before a final measurement of all (data and ancilla) qubits. **c** All syndrome measurements of the k cycles are processed by the decoder. **d** After each cycle, the decoder updates its internal state to represent the most likely set of errors that occurred. **e** After the final measurement, the decoder uses the readout from the data qubits, along with previous syndrome measurements, to declare a final logical state. To this end, the decoder processes the Z-stabilizers obtained directly from the data qubits, finalizing its prediction of most likely errors. The logical parity is then determined as the product of all data qubit parities $\left(\prod_{j=0}^{8} D_j \right)$ once the declared errors are corrected. The logical fidelity $\mathcal{F}_L$ is the probability that this declaration is the same as the initial state ($|0\rangle$)

npj

Density-matrix simulation of small surface
TE O'Brien et al.

6

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

Protocol for measurement of logical error rates. As the performance measure of Surface-17, we study the fidelity of the logical qubit as a quantum memory. We describe our protocol with an example 'run' in Fig. 6. We initialize all qubits in $|0\rangle$ and perform $k = 1, 2, …, 20$ QEC cycles [Fig. 6b]. Although this initial state is not a stabilizer eigenstate, the first QEC cycle projects the system into one of the 16 overlapping eigenstates within the +1 eigenspace for Z stabilizers, which form the logical $|0\rangle$ state.[10] This implies that, in the absence of errors, the first measurement of the Z stabilizers will be +1, whilst that of the X stabilizers will be random. In the following cycles, ancilla measurements of each run [Fig. 6c] are processed using a classical decoding algorithm. The decoder computes a Pauli update after each QEC cycle [Fig. 6d]. This is a best estimate of the Pauli operators that must be applied to the data qubits to transform the logical qubit back to the logical $|0\rangle$ state. The run ends with a final measurement of all data qubits in the computational basis. From this 9-bit outcome, a logical measurement result is declared [Fig. 6e]. First, the four Z-type parities are calculated from the 9 data-qubit measurement outcomes and presented to the decoder as a final set of parity measurements. This ensures that the final computed Pauli update will transform the measurement results into a set that measures +1 for all Z stabilizers. This results in one of 32 final measurements, from which the value of a logical Z operator can be calculated to give the measurement result (any choice of logical operator gives the same result). The logical fidelity $\mathcal{F}_L[k]$ after k QEC cycles is defined as the probability of this declared result matching the initial +1 state.

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

$$\mathcal{F}_L[k] = 1 - \sum_{l\,\text{odd}} \binom{k}{l} \epsilon_L^l (1 - \epsilon_L)^{k-l}. \tag{4}$$

Here, the combinatorial factor counts the number of combinations of l errors in k rounds, given an ϵ_L chance of error per round. This can be simplified to

$$
\begin{aligned}
\mathcal{F}_L[k] &= 1 - \tfrac{1}{2}\sum_l \binom{k}{l} \epsilon_L^l (1-\epsilon_L)^{k-l}\left(1 - (-1)^l\right) \\
&= 1 - \tfrac{1}{2}\left[(1 - \epsilon_L + \epsilon_L)^k - (1 - \epsilon_L - \epsilon_L)^k\right] \\
&= \tfrac{1}{2}\left[1 + (1 - 2\epsilon_L)^k\right].
\end{aligned}
\tag{5}
$$

However, at small k, the decay is dominated by the majority vote, for which $\epsilon_L \propto (k\epsilon_{\text{phys}})^{(d+1)/2}$. For example, for all the Surface-17 decay curves, we observe a quadratic error rate at small k, as opposed to the linear slope predicted by Eq. (5). In order to correct for this, we shift the above equation in k by a free parameter k_0, resulting in Eq. (2). This function fits well to data with $k \geq 3$ in all plots, and thus allows accurate determination of ϵ_L.

The quantumsim simulation package. Quantumsim performs calculations on density matrices utilizing a graphics processing unit in a standard desktop computer. Ancillas are measured at the end of each cycle, and thus not entangled with the rest of the system. As such, it is possible to obtain the effect of the QEC cycle on the system without explicitly representing the density matrix of all 17 qubits simultaneously. The simulation is set up as follows: the density matrix of the nine data qubits is allocated in memory with all qubits initialized to $|0\rangle$. One-qubit and two-qubit gates are applied to the density matrix as completely positive, trace

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

Qubit measurements are simulated as projective and following the Born rule, with projection probabilities given by the squared overlap of the input state with the measurement basis states. In order to capture empirical measurement errors, we implement a black-box measurement model (see Error Models) by sandwiching the measurement between idling processes. The measurement projects the system to a product state of the ancilla and the projected sub-block of the density matrix. We can therefore remove the ancilla from the density matrix and only store its state right after projection, and continue the calculation with the partial density matrix of the other qubits. Making use of the specific arrangement of the interactions between ancillas and data qubits in Surface-17, it is possible to apply all operations to the density matrix in such an order (shown in Supplemental Material) that the total size of the density matrix never exceeds $2^{10} \times 2^{10}$ (nine data qubits plus one ancilla), which allows relatively fast simulation. We emphasize that with the choice of error model in this work, this approach gives the same result as a full simulation on a 17-qubit density matrix. Only the introduction of residual entangling interactions between data and ancilla qubits (which we do not consider in this work) would make the latter necessary. On our hardware (See Supplemental Material), simulating one QEC cycle of Surface-17 with quantumsim takes 25 ms.

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

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

A second highlight of quantumsim is the possibility to quantify the sub-optimality of the decoder. The fidelity of the logical qubit obtained in these numerical simulations is a combination of the error rates of the physical qubits and the approximations made by the decoder. Full density-matrix simulations make it possible to disentangle these two contributions. Namely, the fidelity is obtained by assigning correctness to each of the 512 possible readouts according to 16 outputs of the decoder, and summing the corresponding probabilities accordingly. If the probabilities are known, it is easy to determine the 16 results that a decoder should output in order to maximize fidelity (i.e., the output of the best-possible decoder). This allows placing a decoder UB $\mathcal{F}_L^{\text{max}}$ on logical fidelity as limited by the physical qubits independent of the decoder. Conversely, it also allows quantifying sub-optimality in the decoder used. In fact, we can make the following reverse statement: if our measurement model did not include a declaration error, then we could use the simulation to find the final density matrix of the system conditioned on a syndrome measurement. From this, the simulation could output exactly the 16 results that give $\mathcal{F}_L^{\text{max}}$, so that quantumsim could be used as a maximum-likelihood decoder. In this situation, $\mathcal{F}_L^{\text{max}}$ would not only be an UB, but indeed the performance of the best-possible decoder. However, as we add the declaration errors after simulation, we can only refer to $\mathcal{F}_L^{\text{max}}$ as the decoder UB.

Error models

We now describe the error model used in the simulations. Our motivation for the development of this error model is to provide a limited number of free parameters to study, whilst remaining as close to known experimental data as possible. As such, we have taken well-established theoretical models as a base, and used experimental tomography to provide fixed parameters for observed noise beyond these models. The parameters of the error model are provided in Supplemental Material.

Density-matrix simulation of small surface
TE O'Brien et al.

npj

7

Idling qubits. While idling for a time τ, a transmon in $|1\rangle$ can relax to $|0\rangle$. Furthermore, a transmon in superposition can acquire random quantum phase shifts between $|0\rangle$ and $|1\rangle$ due to $1/f$ noise sources (e.g., flux noise) and broadband ones (e.g., photon shot noise[35] and quasiparticle tunneling[36]). These combined effects can be parametrized by probabilities $p_1 = \exp(-\tau/T_1)$ for relaxation, and $p_\phi = \exp(-\tau/T_\phi)$ for pure dephasing. The combined effects of relaxation and pure dephasing lead to decay of the off-diagonal elements of the qubit density matrix. We model dephasing from broadband sources in this way, taking for T_ϕ the value extracted from the decay time T_2 of standard echo experiments:

$$\frac{1}{T_2} = \frac{1}{T_\phi} + \frac{1}{2T_1}. \tag{6}$$

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

Dephasing from photon noise. The dominant broadband dephasing source is the shot noise due to photons in the readout resonator. This dephasing is present whenever the coupled qubit is brought into superposition before the readout resonator has returned to the vacuum state following the last measurement. This leads to an additional, time-dependent pure dephasing (rates given in Supplemental Material).

One-qubit Y rotations. We model y-axis rotations as instantaneous rotations sandwiched by idling periods of duration $\tau_{g,1Q}/2$. The errors in the instantaneous gates are modeled from process matrices measured by gate-set tomography[30, 37] in a recent experiment.[20] In this experiment, the GST analysis of single-qubit gates also showed that the errors can mostly be attributed to Markovian noise. For simplicity, we thus model these errors as Markovian.

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

C-Z gate error. The C-Z gate is achieved by flux pulsing a transmon into the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing with another, where the 2 denotes the second-excited state of the fluxed transmon. Holding the transmons here for $\tau_{g,2Q}$ causes the probability amplitudes of $|01\rangle$ and $|11\rangle$ to acquire phases.[38] Careful tuning allows the phase ϕ_{01} acquired by $|01\rangle$ (the single-qubit phase ϕ_{1Q}) to be an even multiple of 2π, and the phase ϕ_{11} acquired by $|11\rangle$ to be π extra. This extra phase acquired by $|11\rangle$ is the two-qubit phase ϕ_{2Q}. Single-qubit and two-qubit phases are affected by flux noise because the qubit is first-order sensitive during the gate. Previously, we discussed the single-qubit phase error. In Supplemental Material, we calculate the corresponding two-qubit phase error $\delta\phi_{2Q}$. Our full (but simplistic) model of the C-Z gate consists of an instantaneous C-Z gate with single-qubit phase error $\delta\phi_{1Q}$ and two-qubit phase error $\delta\phi_{2Q} = \delta\phi_{1Q}/2$, sandwiched by idling intervals of duration $\tau_{g,2Q}/2$.

Measurement. We model qubit measurement with a black-box description using parameters obtained from experiment. This description consists of the eight probabilities for transitions from an input state $|i\rangle \in \{|0\rangle, |1\rangle\}$ into pairs $(m, |o\rangle)$ of measurement outcome $m \in \{+1, -1\}$ and final state $o \in \{|0\rangle, |1\rangle\}$. By final state we mean the qubit state following the photon-depletion period. Input superposition states in the computational bases are first projected to $|0\rangle$ and $|1\rangle$ following the Born rule. The probability tree (the butterfly) is then used to obtain an output pair $(m, |o\rangle)$. These experimental parameters can be described by a six-parameter model (described in detail in Supplemental Material), consisting of periods of enhanced noise before and after a point at which the qubit is perfectly projected, and two probabilities $\epsilon_{RO}^{|i\rangle}$ for wrongly declaring the result of this projective measurement. In Supplemental Material, a scheme for measuring these butterfly parameters and mapping them to the six-parameter model is described. In experiment, we find that the readout errors $\epsilon_{RO}^{|i\rangle}$ are almost independent of the qubit state $|i\rangle$, and so we describe them with a single readout error parameter ϵ_{RO} in this work.

Data availability
The data that support the plots within this paper and other findings of this study are available from the corresponding author (L.DiCarlo@tudelft.nl) upon request.

Code availability
The computer code that was use to generate these results is available from the corresponding author (L.DiCarlo@tudelft.nl) upon request.

AUTHOR CONTRIBUTIONS
T.E.O. wrote the decoder algorithm, and calculated the analytic extension to Surface-49. B.T. wrote the quantumsim program used for all density matrix simulations, and calculated simulation parameters from the underlying model. L.D. provided the underlying model parameters, and calculated simulation parameters from the underlying model. All authors contributed equally to the writing of the paper.

ADDITIONAL INFORMATION
Supplementary Information accompanies the paper on the *npj Quantum Information* website (https://doi.org/10.1038/s41534-017-0039-x).

Competing interests: The authors declare that they have no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1. O'Malley, P. J. J. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016).
2. Barends, R. et al. Digital quantum simulation of fermionic models with a super-conducting circuit. *Nat. Commun.* **6**, 7654 (2015).
3. Langford, N. K. et al. Experimentally simulating the dynamics of quantum light and matter at ultrastrong coupling. Preprint at http://arxiv.org/abs/1610.10065 (2016).
4. Kelly, J. et al. State preservation by repetitive error detection in a super-conducting quantum circuit. *Nature* **519**, 66–69 (2015).
5. Ristè, D. et al. Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nat. Commun.* **6**, 6983 (2015).
6. Córcoles, A. D. et al. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Commun.* **6**, 6979 (2015).
7. Ofek, N. et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441 (2016).
8. Boixo, S. et al. Characterizing quantum supremacy in near-term devices. Preprint at https://arxiv.org/abs/1608.00263 (2016).
9. Martinis, J. M. Qubit metrology for building a fault-tolerant quantum computer. *npj Quantum Inf.* **1**, 15005 (2015).
10. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A.* **86**, 032324 (2012).
11. Landahl, A. J., Anderson, J. T. & Rice, P. R. Fault-tolerant quantum computing with color codes. Preprint at http://arxiv.org/abs/1108.5738 (2011).
12. Yoder, T. J. & Kim, I. H. The surface code with a twist. Preprint at http://arxiv.org/abs/1612.04795 (2016).

np|

Density-matrix simulation of small surface
TE O'Brien et al.

8

13. Tomita, Y. & Svore, K. M. Low-distance surface codes under realistic quantum noise. *Phys. Rev. A* **90**, 062320 (2014).
14. Fowler, A. G., Stephens, A. M. & Groszkowski, P. High-threshold universal quantum computation on the surface code. *Phys. Rev. A* **80**, 052312 (2009).
15. Heim, B., Svore, K. M. & Hastings, M. B. Optimal circuit-level decoding for surface codes. Preprint at http://arxiv.org/abs/1609.06373 (2016).
16. Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
17. Horsman, C., Fowler, A. G., Devitt, S. & Meter, R. V. Surface code quantum computing by lattice surgery. *New J. Phys.* **14**, 123011 (2012).
18. Versluis, R. et al. Scalable quantum circuit and control for a superconducting surface code. Preprint at http://arxiv.org/abs/1612.08208 (2016).
19. Bultink, C. C. et al. Active resonator reset in the nonlinear dispersive regime of circuit qed. *Phys. Rev. Appl* **6**, 034008 (2016).
20. Rol, M. A. et al. Restless tuneup of high-fidelity qubit gates. *Phys. Rev. Appl.* **7**, 041001 (2017).
21. Asaad, S. et al. Independent, extensible control of same-frequency superconducting qubits by selective broadcasting. *npj Quantum Inf.* **2**, 16029 (2016).
22. Walter, T. et al. Realizing Rapid, High-Fidelity, Single-Shot Dispersive Readout of Superconducting Qubits. Preprint at http://arxiv.org/abs/1701.06933 (2017).
23. Frisk Kockum, A., Tornberg, L. & Johansson, G. Undoing measurement-induced dephasing in circuit QED. *Phys. Rev. A* **85**, 052318 (2012).
24. Paik, H. et al. Experimental demonstration of a resonator-induced phase gate in a multiqubit circuit-qed system. *Phys. Rev. Lett.* **117**, 250502 (2016).
25. Gustavsson, S. et al. Suppressing relaxation in superconducting qubits by quasiparticle pumping. *Science* **354**, 1573–1577 (2016).
26. Terhal, B. M. Quantum error correction for quantum memories. *Rev. Mod. Phys.* **87**, 307–346 (2015).
27. Fowler, A. G., Sank, D., Kelly, J., Barends, R. & Martinis, J. M. Scalable extraction of error models from the output of error detection circuits. Preprint at https://arxiv.org/abs/1405.1454 (2014).
28. Delfosse, N. & Tillich, J. P. A decoding algorithm for css codes using the x/z correlations. In *2014 IEEE International Symposium on Information Theory*, 1071–1075 (2014).
29. Fowler, A. G. Optimal complexity correction of correlated errors in the surface code. Preprint at https://arxiv.org/abs/1310.0863 (2013).
30. Blume-Kohout, R. et al. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. Preprint at https://arxiv.org/abs/1310.4492 (2013).
31. Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500 (2014).
32. Chen, Z. et al. Multi-photon sideband transitions in an ultrastrongly-coupled circuit quantum electrodynamics system. Preprint at https://arxiv.org/abs/1602.01584 (2016).
33. Fowler, A. G. Coping with qubit leakage in topological codes. *Phys. Rev. A* **88**, 042308 (2013).
34. Gottesman, D. *Stabilizer Codes and Quantum Error Correction*. Ph.D. thesis Caltech, (1997).
35. Sears, A. P. et al. Photon shot noise dephasing in the strong-dispersive limit of circuit QED. *Phys. Rev. B* **86**, 180504 (2012).
36. Ristè, D. et al. Millisecond charge-parity fluctuations and induced decoherence in a superconducting transmon qubit. *Nat. Commun.* **4**, 1913 (2013).
37. Blume-Kohout, R. et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.* **8**, 14485 (2017).
38. DiCarlo, L. et al. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240 (2009).