

Overcoming leakage in quantum error correction

In the format provided by the
authors and unedited

Supplementary information for *Overcoming leakage in quantum error correction*

CONTENTS

S1. Effects of leakage on diabatic CZ gate	1
S2. Leakage Removal Strategy Details	1
S3. Effect of reset strategies on leakage dynamics	3
S4. Effect of reset strategies on QEC error detection	4
S5. Fitting techniques for logical error probability	5
S6. Surface code simulations well below threshold	6
S7. Error correlations in the surface code experiment	6
References	8

S1. EFFECTS OF LEAKAGE ON DIABATIC CZ GATE

During the diabatic CZ gate, additional levels are placed on resonance and contribute to the leakage transport phenomenon depicted in the main text Figure 2. The resonance $|30\rangle \leftrightarrow |12\rangle$ allows a two-photon transition mediated by $|21\rangle$, which is detuned by around the nonlinearity η . If g is the induced coupling between $|11\rangle$ and $|20\rangle$, then the effective couplings are:

$$\begin{aligned}
 g_{|30\rangle \leftrightarrow |21\rangle} &= \sqrt{3}g, \\
 g_{|21\rangle \leftrightarrow |12\rangle} &= 2g, \\
 g_{\text{eff}} = g_{|30\rangle \leftrightarrow |12\rangle} &= -g_{|21\rangle \leftrightarrow |12\rangle} \times g_{|30\rangle \leftrightarrow |21\rangle} / \eta.
 \end{aligned}$$

Then, for a CZ gate where g is maintained for time t , the population transport P_t for $|30\rangle \leftrightarrow |12\rangle$ can be estimated as

$$P_t = \sin^2(g_{\text{eff}}t).$$

To measure the leakage transport in the diabatic CZ gate, we calibrate a readout pulse capable of distinguishing all of the four lowest qubit energy levels, as shown in Figure S1a. When we encounter $|4\rangle$ during measurement using this readout pulse, we assign and identify it as $|3\rangle$. The baseline experiment consists of preparing a given two-qubit state using microwave drives and then performing simultaneous readout of the qubits. The “Baseline” matrix of Figure S1b shows the results of this experiment, illustrating that performing readout simultaneously does not impact the high distinguishability between all two-qubit states. We can also see that the majority of the error in this simultaneous readout is due to T_1 decay during the readout process. These decay channels reduce the

populations on the main diagonal by a few percent, and become more prominent for higher levels. The “With CZ gate” matrix of Figure S1b shows the same experiment with a CZ gate inserted between state preparation and measurement. We then see new off-diagonal processes corresponding to leakage transport. We subtract the “Baseline” matrix from the “With CZ gate” matrix to produce the matrix shown in Figure 2b of the main text. The leakage transport processes that are most relevant to the spread of leakage in our system correspond to the $|30\rangle \leftrightarrow |12\rangle$ and $|31\rangle \leftrightarrow |22\rangle$ rotations. In Figure S1c, we numerically show the individual relative population transport values ΔP_t for these two processes. We take the mean of the absolute value of the relative population transport values to determine the values at the bottom of Figure 2b of the main text.

To measure the spurious phase shift generated by leaked higher-energy qubits undergoing the CZ gate, we perform the experiment detailed in Figure S2a. The higher-energy qubit in the pair is prepared in each of $|0\rangle$, $|1\rangle$, and $|2\rangle$, while a Ramsey experiment is performed on the lower-energy qubit with an interleaved CZ gate. On the lower-energy qubit, we vary the tomography phase ϕ_T of the second pulse relative to the first pulse in the Ramsey experiment to record data as shown in Figure S2b for each input state. We fit a sinusoid to the experimental data and extract the phase offset ϕ . We perform this experiment on 20 pairs of qubits, and show the empirical cumulative distribution function of the extracted phase offsets in Figure 2e of the main text.

S2. LEAKAGE REMOVAL STRATEGY DETAILS

We studied three leakage removal strategies; *No reset*, *MLR* and *DQLR*. We now describe them in greater detail.

For *No reset*, we add no additional operations at the end of each cycle. Because this prepares the qubit for the next cycle in whichever state was measured rather than deterministically in $|0\rangle$, this also requires the redefinition of the detectors in the surface and bit-flip code circuits: rather than comparing time-neighboring measurements, we compare time-next-neighboring measurements on the same measure qubit to detect errors [1]. This redefinition has an insignificant impact on code performance, especially when compared to the studied effects of leakage, and so we neglect it from our analysis.

For *MLR*, we add the multi-level reset (MLR) operation introduced in [2] on the measure qubits at the end of each cycle. Additional pulse shaping on the diabatic return and calibration improvements allow us to achieve

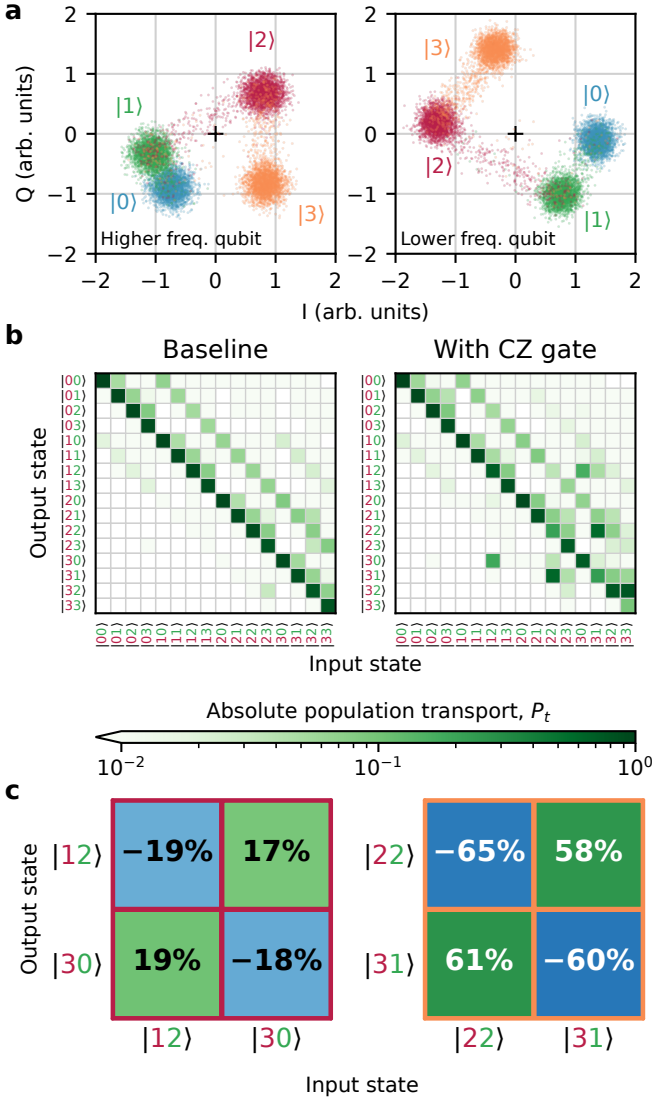


Figure S1. **Measuring leakage transport.** a) Readout clouds for the pair of qubits used in the leakage transport experiment, showing good distinguishability between all of the lowest four qubit energy states. b) Raw measured population transport matrices for the two transport experiments. In both matrices, we can see the effect of T_1 decay during measurement, which is enhanced for the higher levels. Subtracting “Baseline” from “With CZ gate” produces the matrix shown in Figure 2b of the main text. c) Individual relative population transport values ΔP_t for relevant leakage transport mechanisms $|30\rangle \leftrightarrow |12\rangle$ and $|31\rangle \leftrightarrow |22\rangle$. The average of the norm of these values produce the population transport values at the bottom of Figure 2b of the main text.

gate times of 160 ns without impacting performance.

For *DQLR*, we first perform the MLR operation on all measure qubits, as in *MLR*. Following that, we perform the *DQLR* operation, which consists of a *LeakageISWAP* gate between pairs of measure and data qubits, and a second reset of the measure qubits. The *LeakageISWAP* gate is similar to the diabatic CZ gate used in the sur-

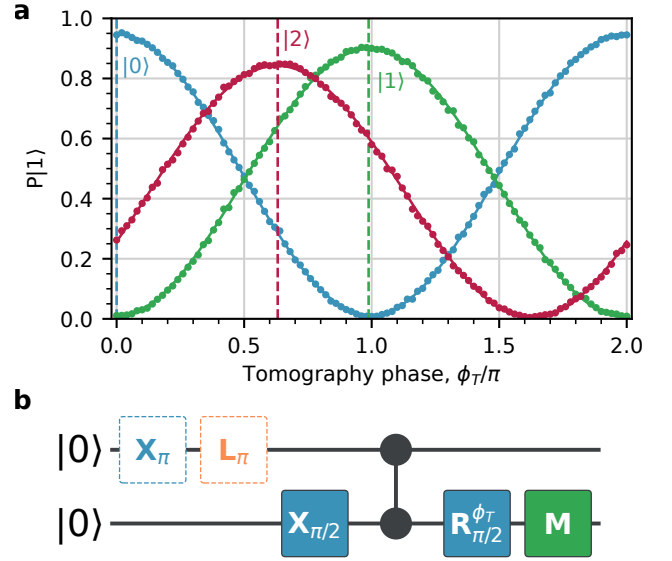


Figure S2. **Measuring leakage phases.** a) Raw data for the leakage phase experiment on a single pair of qubits. Each of $|0\rangle$, $|1\rangle$, and $|2\rangle$ is prepared on the higher-energy qubit, and then a Ramsey experiment with an interleaved CZ gate is performed on the lower-energy qubit. The solid lines are sinusoidal fits and the dashed lines indicate the extracted phase shifts ϕ shown in the main text. b) The hardware circuit executed in the above experiment in the $|2\rangle$ case. The two initial rotations (X and L) on the higher-energy qubit (top) are removed when preparing $|0\rangle$, and the second rotation L is removed when preparing $|1\rangle$.

face code and bit-flip code cycles, but executes an *ISWAP* gate in the $|11\rangle - |20\rangle$ subspace. Specifically, this entails performing the total-excitation-preserving swap $|20\rangle \rightarrow i|11\rangle$. The second reset gate differs from the *MLR* reset gate in that it can be optimised for only extracting a single photon, and so can be made significantly faster. We utilize a Landau-Zener-Stückelberg (LZS) double passage between the qubit and readout resonator to perform this fast reset, and dub the gate *LZS reset*. We note that this *DQLR* operation relies on the high fidelity of the preceding *MLR* gate; when the measure qubit is prepared in $|0\rangle$, the *LeakageISWAP* gate removes $|2\rangle$ on the data qubits. Any reset error leaving $|1\rangle$ on the measure qubits after the *MLR* gate will be converted into leakage on the data qubit by the *LeakageISWAP* gate. Our results show that this error path is sufficiently low in probability so as not to increase the leakage population on the data qubits.

Ideally, *DQLR* should not induce additional errors on the data qubit. In particular, when the lower-energy measure qubit is in $|0\rangle$, the *LeakageISWAP* gate should act as an identity operation on the data qubit computational basis. However, the non-zero time to execute the *DQLR* operation introduces incoherent errors caused by relaxation, in addition to coherent errors from miscalibration. We evaluate the impact of the *DQLR* operation

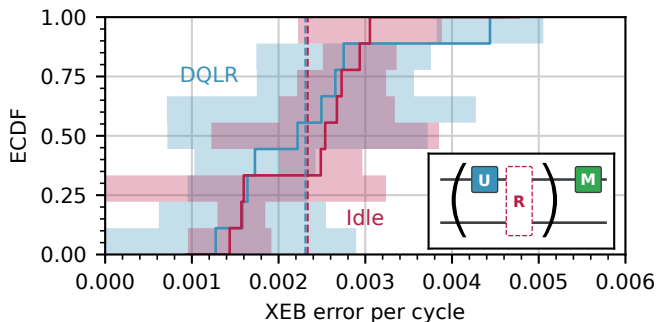


Figure S3. **Cross-entropy benchmarking (XEB) of DQLR operation.** Inferred XEB error per cycle for 9 data qubits when the DQLR operation is applied on the data qubit, compared to XEB error per cycle when the data qubit idles for the equivalent time. Shaded region indicates 1 SD error of inferred XEB error per cycle. Vertical dashed lines indicate the mean XEB error per cycle over all data qubits. (Inset) XEB circuit where random single-qubit unitary rotations (**U**) are repeatedly applied to the target qubit, interleaved with a reset operation (**R**), which is either the DQLR operation (**DQLR**) or idle (**Idle**). The final state of the target qubit is measured. The cross-entropy between the measured and expected distribution of states is calculated and the resulting XEB error per cycle is inferred.

on the data qubit state using cross-entropy benchmarking (XEB). The inset of Figure S3 shows the experimental circuit used to evaluate XEB error. The upper and lower qubits mimic the role of data and measure qubits, respectively. The section of the circuit within the parentheses is repeated a variable number of times, and the final state of the upper qubit is measured. The cross-entropy of the measured and expected distributions of states is calculated as a function of repetitions, and then the XEB error per repetition is extracted. In a given repetition, a random unitary **U** is executed on the upper qubit, followed by a reset operation **R**. In the case of **DQLR**, **R** is substituted with the DQLR operation. We compare this to the **Idle** case, where **R** is replaced with waiting for the duration of the DQLR operation. We carry this measurement out over the 9 pairs of data and measure qubits corresponding to the pairings used in the distance-3 surface code experiment. By comparing the resulting distribution of XEB error per cycle for **DQLR** and **Idle**, we conclude that the DQLR operation does not induce significantly more errors than idling for the equivalent duration. Furthermore, the mean error rate of less than 2.5×10^{-3} per cycle when using the DQLR operation is low enough so that the operation is suitable to be added to a sensitive QEC circuit such as the surface code. We note that leakage is still a relevant consideration in this XEB experiment, and is captured by XEB error per cycle as an incoherent error. Thus, it is possible that the leakage removal properties of **DQLR** result in underreported XEB error per cycle when compared

to **Idle**, which allows for leakage to accumulate over the course of the XEB circuit.

In Table S1, we provide additional error and timing components of the DQLR operation. The combination of speed, high leakage removal fidelity, and low computational basis error underlies the performant nature of the operation in a QEC setting. This allows us to append the operation at the end of each QEC cycle while also observing improvements in logical performance.

S3. EFFECT OF RESET STRATEGIES ON LEAKAGE DYNAMICS

As we have demonstrated in the main text, leakage transport can move leakage population from qubit to qubit in a structured circuit such as the surface code. Once a data qubit is leaked, leakage removal strategies must be employed or the leakage population may remain for many QEC cycles and cause additional leakage and leakage-induced error through leakage transport. In Figure S4, we evaluate the dynamics of leakage population in a surface code under the three leakage removal strategies discussed in this work. We fully inject leakage at the beginning of the first cycle by performing a $|1\rangle \rightarrow |2\rangle$ rotation on the central data qubit to obtain near-50% population of $|2\rangle$. We measure excess leakage population, which is defined as the leakage population without injection subtracted from the leakage population with injection. This allows us to separate the contribution of intrinsic heating to leakage population dynamics from the injected leakage population.

For *No reset*, leakage transport allows for leakage to move freely throughout the qubits surrounding the central data qubit, eventually resulting in measurable excess leakage population in nearly all of the 17 qubits involved in the distance-3 surface code. The lack of leakage removal procedures on either the measure qubits or data qubits leaves T_1 energy relaxation of $|2\rangle$ as the primary mechanism for leakage population mitigation. As we showed in Figure 1c of the main text, T_1 of $|2\rangle$ can be longer than 10 surface code cycles, and this number is expected to increase as qubit coherence improves and as surface code cycle durations shorten. Hence, relying on energy relaxation is not a viable strategy for leakage removal in QEC. This is partially addressed by **MLR** by

Table S1. **DQLR operation error and timing components.**

Parameter	Fidelity (Mean)		Gate Time
Multi-level reset	$P(1\rangle \rightarrow 0\rangle)$	99.86%	160 ns
LZS reset	$P(1\rangle \rightarrow 0\rangle)$	99.57%	25 ns
LeakageISWAP	$P(11\rangle \rightarrow i 20\rangle)$	98.9%	28 ns

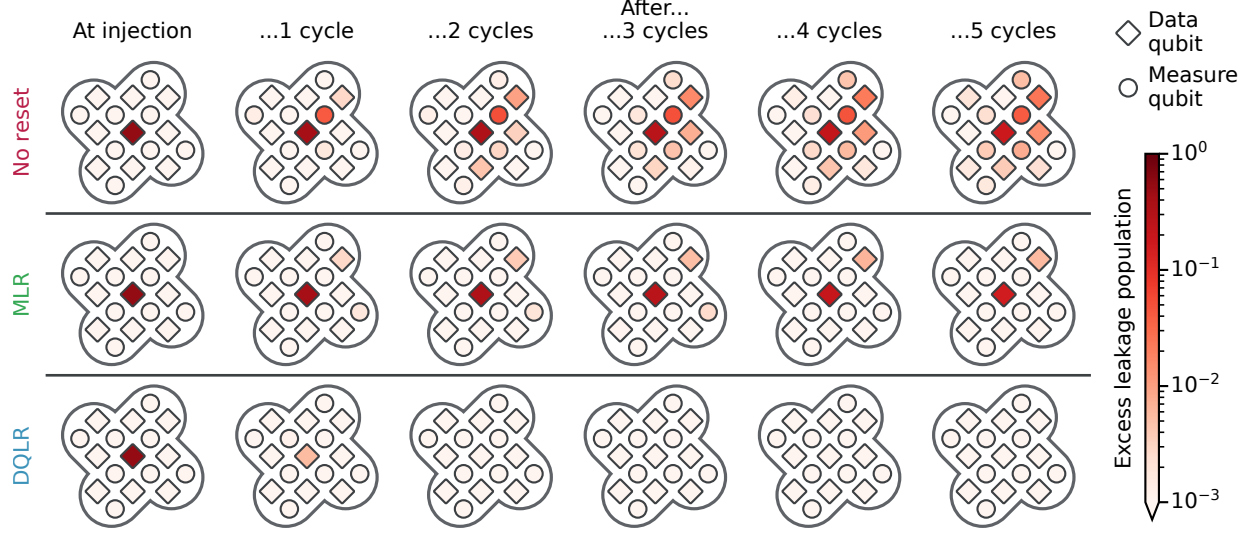


Figure S4. **Leakage dynamics in a surface code with different removal strategies.** Comparison of excess leakage population dynamics over 5 cycles for all qubits in a distance-3 surface code after a full $|1\rangle \rightarrow |2\rangle$ leakage injection during the first cycle. With *No reset*, leakage transport mechanisms lead to increasing leakage population over nearly all qubits involved in the code. With the introduction of *MLR*, a sink for leakage population is present on measure qubits, mitigating the spread of leakage from leakage transport effects. The central leakage-injected data qubit still remains significantly leaked, even after 5 cycles. Using *DQLR*, leakage populations on all qubits in the code are brought to about 1×10^{-3} or lower within 2 cycles.

applying a multi-level reset gate on all measure qubits at the end of each cycle. In Figure S4, the effect of this operation appears as reduced excess leakage populations on all measure qubits at the end of every cycle. However, leakage transport within a QEC cycle can still move leakage population beyond nearest-neighbor qubits. This is readily observed in the excess leakage population of the data qubit two sites away from the central data qubit, which exhibits increasing population even though the measure qubit between the two data qubits has its leakage population removed by the MLR operation at the end of every cycle.

DQLR suppresses the ability of leakage to hop between data qubits by directly removing a large fraction of all the data qubits' leakage populations at the end of each cycle. In particular, the central data qubit has its excess leakage population reduced to less than 1% after the first cycle. Similarly, other data qubits that previously leaked due to leakage transport have their excess leakage populations reduced close to the measurement floor. The shortened lifetime of leakage on all qubits is clearly seen for *DQLR* after two QEC cycles, where excess leakage population over all qubits is less than 1×10^{-3} .

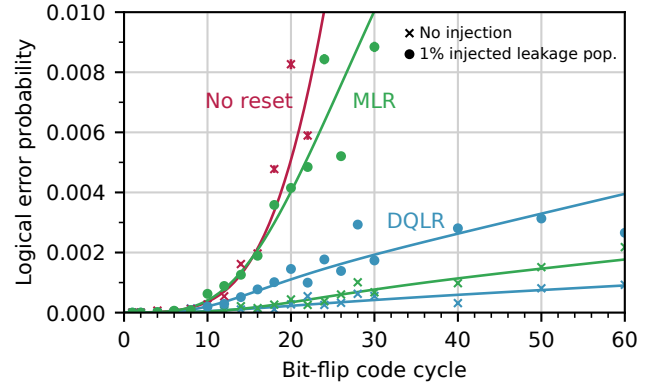


Figure S5. **Logical performance of bit-flip code in time.** Logical error probability of the distance-21 bit-flip code over 60 cycles for the three reset strategies *No reset* (red), *MLR* (green), and *DQLR* (blue). For *MLR* and *DQLR*, we also execute the code while injecting 1% leakage population per cycle. Early in the code (fewer than 5 cycles), boundary effects cause all three reset strategies to perform similarly. As the code progresses through more and more cycles, however, logical performance for the three strategies diverge as leakage populations are handled differently.

S4. EFFECT OF RESET STRATEGIES ON QEC ERROR DETECTION

In Figure S5, we compare the time dynamics of the bit-flip code under the three different reset strategies pre-

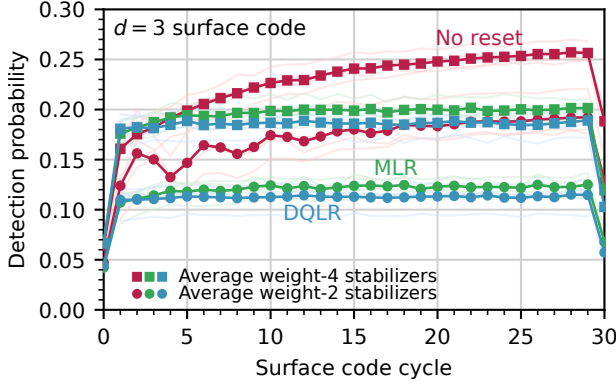


Figure S6. **Distance-3 surface code detection probabilities.** Average detection probabilities over 30 surface code cycles for weight-2 (circles) and weight-4 (squares) stabilizers for three leakage removal strategies *No reset* (red), *MLR* (green), and *DQLR* (blue). Individual stabilizer detection probabilities are shown as lighter lines.

sented in the main text. We execute a distance-21 bit-flip code over 60 cycles, as described by the circuit in Figure 4d of the main text. For *MLR* and *DQLR*, we inject both 0% and 1% leakage population in each cycle, whereas for *No reset* we do not inject leakage. In the first few cycles of code execution, differences in the logical performance of the bit-flip code between the various leakage removal strategies and injection populations are difficult to distinguish – we attribute this to time-boundary effects where physical errors have not sufficiently accumulated to manifest as a logical error, and statistical limitations where the logical error probability is much smaller than what is resolvable by the number of trials. However, the logical performance for the three leakage removal strategies begins to diverge after about 10 cycles. The accumulation of leakage on measure and data qubits causes a rapid rise of logical error probability for *No reset*, where it exceeds 1×10^{-2} error by 25 cycles. This is in contrast to *MLR* and *DQLR* without leakage injection, which continue to have less than 3×10^{-3} logical error probability through 60 cycles; *DQLR* has the best performance with about 1×10^{-3} logical error probability at 60 cycles.

Turning to the cases where we inject 1% leakage population per cycle for *MLR* and *DQLR*, we observe a notable distinction in logical error probability scaling over cycle number. With *MLR*, the logical error probability rises to about 1×10^{-2} by 30 cycles, with a similar qualitative behavior to *No reset* without leakage injection. However, when we use *DQLR*, the code sustains logical error probabilities of less than 5×10^{-3} over 60 cycles. This is a signature that QEC scaling in time can be more easily achieved when using *DQLR*.

In Figure S6, we present the average weight-2 stabilizer detection probabilities in addition to the average weight-4 stabilizer values already shown in Figure 5a of the main

text. Additionally, we show the detection probabilities associated with the individual stabilizers. We can draw parallel conclusions for the behavior of weight-2 stabilizers as we did for weight-4 stabilizers in the main text. For *No reset*, the average weight-2 stabilizer detection probabilities rise and do not stabilize over the course of 30 surface code cycles, and even exhibit damped oscillations at early cycles. When using *MLR*, the detection probability stability improves significantly and the average probability only rises by about 1×10^{-2} over 10 cycles before flattening. However, the best performance and stability is achieved with *DQLR*, where the average weight-2 stabilizer detection probability remains at 11% over the course of the entire 30-cycle experiment.

S5. FITTING TECHNIQUES FOR LOGICAL ERROR PROBABILITY

We employ phenomenological models to fit experimental and simulated data. The models are implemented on the logical error per cycle ε of the QEC code. In experimental and simulated data, we measure logical error probability p_L after n QEC cycles. The conversion between ε and p_L after n cycles is given by

$$\varepsilon = \frac{1 - (1 - 2p_L)^{\frac{1}{n}}}{2}, \quad (\text{S1})$$

$$p_L = \frac{1 - (1 - 2\varepsilon)^n}{2}. \quad (\text{S2})$$

With respect to injected error population P , ε can be modeled as a power law with an offset,

$$\varepsilon(P) = a(P + P_0)^b, \quad (\text{S3})$$

where a , b , P_0 are phenomenological free parameters. We do not ascribe a physical meaning to P_0 even though it may appear to be “intrinsic” error at $P = 0$. We use this power law model to fit the experimental data in Figures 4a and 5b of the main text, as well as the numerical simulations in Figure 4a of the main text.

In order to model $1/\Lambda_{5/7}$ as a function of injected leakage population P_L in Figure 5c of the main text, we take the inverse of the ratio between distance-5 and distance-7 logical error per cycle,

$$\Lambda_{5/7}(P_L) = \frac{\varepsilon_5(P_L)}{\varepsilon_7(P_L)}. \quad (\text{S4})$$

For logical error probability p_L with respect to cycle n of a bit-flip code operated well below threshold and in the presence of leakage dynamics (Figure S5), we use a Gompertz model to describe logical error rate ε ,

$$\varepsilon(n) = a \exp(-b \exp(-cn)), \quad (\text{S5})$$

where a , b , and c are phenomenological free parameters. A Gompertz model can partly capture the transient dynamics of ε at small n , where time-boundary effects and still-increasing leakage populations make ε highly dependent on n . At large n , ε tends to a constant as leakage populations stabilize.

S6. SURFACE CODE SIMULATIONS WELL BELOW THRESHOLD

To gain insight into the future importance of leakage removal for scaling quantum error correction, we performed numerical simulations of distance-5 and distance-7 surface codes and evaluated their logical performance subject to different levels of leakage injection carried out using the circuit shown in the inset of Figure 5b of the main text. We performed these simulations using a Kraus operator simulation detailed in [3]. We include operators that accurately reflect leakage transport, leakage phase errors, and the leakage removal parameters for both the *MLR* and *DQLR* strategies, but do not include any sources of leakage in the baseline error model at zero leakage injection. The error model for the baseline simulations is detailed in Table S2. We repeat these simulations with varying amounts of injected leakage population under two leakage removal strategies, *MLR* and *DQLR*, which is presented in the Figure 5c of the main text.

We fit the data from the numerical simulations in Figure 5c of the main text with a power law ratio model (Equation S4). Additionally, we fit the data for *DQLR* with a line. The high degree of linearity ($R^2 = 0.983$, $1/\Lambda_{5/7} \approx 111 \times P_L + 0.2$) suggests that the surface code error budget $1/\Lambda_{5/7}$ for leakage under *DQLR* can be linearized and modeled as an uncorrelated error source.

Table S2. **Hypothetical device error model for simulations carried out in Figure 5c of the main text.**

Parameter	Value
Single qubit gate Pauli error	2×10^{-4}
CZ gate Pauli error	1×10^{-3}
Readout and reset error	1×10^{-2}
Idling Pauli error from relaxation	3×10^{-3}
from dynamical decoupling	1×10^{-3}
Qubit T_1	75 μ s
Qubit T_2	75 μ s
Single qubit gate time	15 ns
CZ gate time	25 ns
Combined readout and reset time	300 ns

S7. ERROR CORRELATIONS IN THE SURFACE CODE EXPERIMENT

One technique to evaluate error correlations in the surface code is to employ p_{ij} correlation matrices [2–5]. By analyzing the data presented in Figure S6 with p_{ij} correlation matrices, we can elucidate the presence of error correlations in space and time over the course of the QEC experiment.

For p_{ij} matrices, i and j correspond to nodes in the detection graph, each of which have stabilizer (or *space*, s) and time (t) coordinates; $i = (s, t)$, $j = (s', t')$. To first focus on correlations in time, we average the matrix elements p_{ij} with $s = s'$ over all t and t' , producing autocorrelation matrices $\bar{p}_{t,t'}$, defined as

$$\bar{p}_{t,t'} = \frac{\sum_{s=s'} p_{ij}}{\sum_{s=s'} 1}. \quad (\text{S6})$$

The autocorrelation matrix $\bar{p}_{t,t'}$ reports the average of the probabilities of error graph edges between nodes i and j associated to the same stabilizer s and for arbitrary time separation $|t - t'|$.

In Figure S7a, we show averaged autocorrelation matrices $\bar{p}_{t,t'}$ for both X - and Z -basis stabilizers over the 30 cycle experiment under the three leakage removal strategies investigated in this work. Z -basis stabilizers do not report detection probabilities at the time-boundary cycles 0 and 30 and thus do not have $\bar{p}_{t,t'}$ elements associated with those cycles. Independent Pauli errors present themselves as detection events on consecutive cycles $|t - t'| = 1$. In such an ideal setting with only fully independent errors, we expect non-local correlations to vanish, *i.e.* $\bar{p}_{t,t'} = 0$ where $|t - t'| > 1$. Leakage and other time-non-local error sources will break this assumption and force those $\bar{p}_{t,t'}$ elements to be non-zero.

For *No reset*, non-local correlations immediately appear at cycle 1 and increase in intensity over the course of the 30-cycle experiment. The high values for non-local $\bar{p}_{t,t'}$ matrix elements greater than 1% at large correlation distances $|t - t'|$ indicate that non-local correlations are significant contributors to logical performance degradation, and that QEC cannot be practically scaled in time under these conditions [6].

For *MLR*, non-local correlations are visibly reduced when compared to *No reset*. Still, an impactful degree of non-local correlation remains, primarily stemming from data qubit leakage and resultant leakage-induced correlated errors, such as CZ gate phase shifts highlighted in Figure 2e of the main text.

For *DQLR*, nearly all non-local correlations are heavily suppressed in the experiment. This is seen by very small correlation magnitudes less than 0.2% at correlation distances greater than 1. Qualitatively, the effective suppression of time-non-local correlations suggests

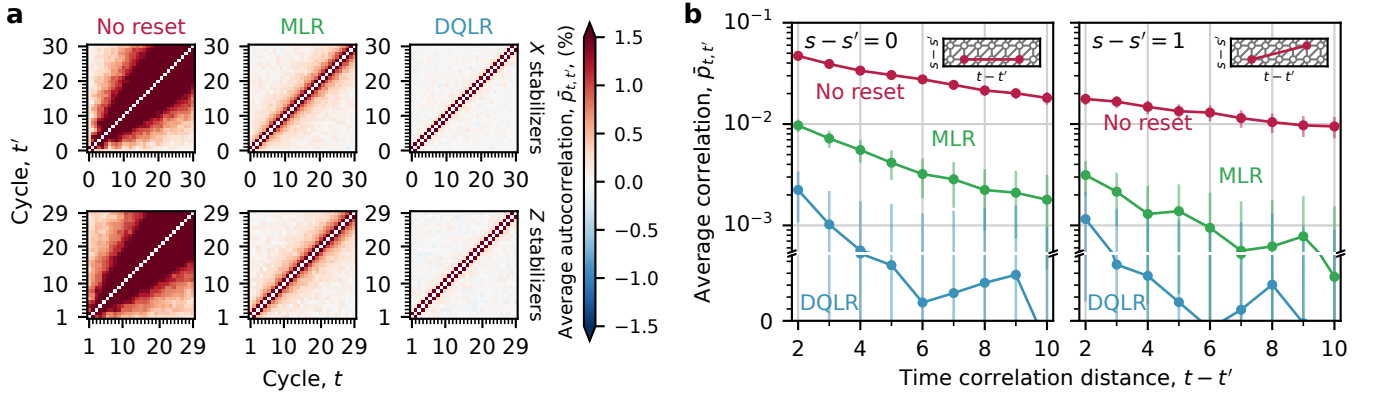


Figure S7. **Surface code p_{ij} correlations.** a) Average autocorrelated p_{ij} matrices $\bar{p}_{t,t'}$ of X and Z stabilizers under different leakage removal strategies. Correlations along the upper and lower diagonals ($p_{t,t'}$ where $|t - t'| = 1$) represent independent timelike errors, whereas non-local correlations ($p_{t,t'}$ where $|t - t'| > 1$) can be primarily attributed to leakage-induced correlated errors. For *No reset*, non-local correlations intensify as the code is executed in time, suggesting increasing leakage-induced correlated errors from growing leakage population. With *MLR*, non-local correlations are reduced but remain, which are manifestations of data qubit leakage-induced correlated errors. For *DQLR*, complete leakage removal over all qubits results in suppression of non-local correlations to about 1×10^{-3} or lower. b) (Left) Averaged autocorrelation $\bar{p}_{t,t'}$; $s - s' = 0$ over all stabilizers for cycles t and t' in 19–29, inclusive, under different leakage removal strategies. Average non-local correlation magnitudes do not exceed 2×10^{-3} for *DQLR*, whereas *No reset* and *MLR* have exponential decays in correlation magnitude with respect to correlation distance. The inset shows an example of a distance-4 long-time error edge on a detection graph. Error bars correspond to 1 SD error. (Right) Averaged nearest-neighbor correlation $\bar{p}_{t,t'}$; $s - s' = 1$ over all stabilizers, extracted under the same conditions as the averaged autocorrelation values. The inset shows an example of a distance-4 long-diagonal error edge on a detection graph. Error bars correspond to 1 SD error.

we are much closer to fulfilling the QEC requirement of uncorrelated errors. Furthermore, this is evidence that our *DQLR* operation does not introduce additional unwanted correlations within the experiment.

We isolate and average $\bar{p}_{t,t'}$ matrix elements for cycles t and t' in 19–29, inclusive, in Figure S7b. This offers a quantitative profile to the degree of non-local correlations present in the surface code as a function of time correlation distance $t - t'$, which correspond to the probability of encountering *long-time error edges* in the detection graph. Again, under ideal circumstances with fully independent errors, the correlation strength $\bar{p}_{t,t'}$ should be 0 at all $t - t'$ greater than 1. *DQLR* most closely approximates this condition, where $|\bar{p}_{t,t'}|$ exceeds 2×10^{-3} only for distance-2 correlation and otherwise is below 0.2% for all distances up through 10. The 1 SD error bars for *DQLR* suggest that we cannot resolve variations in $|\bar{p}_{t,t'}| < 1 \times 10^{-3}$ for these non-local correlations. However, for *MLR*, $|\bar{p}_{t,t'}|$ is about an order of magnitude higher at 1% for distance-2 correlation, and slowly decays with distance, remaining above 0.1% even at distance-10. Finally, *No reset* never has $|\bar{p}_{t,t'}| < 1\%$ for any of the correlation distances considered here.

In order to evaluate correlations across space, we can perform a similar analysis but now with nearest-neighbor

stabilizers,

$$p_{t,t'} = \frac{\sum_{s-s'=1} p_{ij}}{\sum_{s-s'=1} 1}. \quad (\text{S7})$$

The resulting average time correlations, which indicate the probability of encountering a *long-diagonal error edge*, are shown in the right panel of Figure S7b. By only considering time correlation distances of $t - t' > 1$, we primarily exclude contributions from CZ gate errors, which manifest as *diagonal edges* at $s - s' = 1$ and $t - t' = 1$ [3]. The remaining correlations at $t - t' > 1$ then predominantly arise from leakage and crosstalk and should be zero under ideal circumstances without these error sources.

Similar to the average correlation $\bar{p}_{t,t'}$ for $s - s' = 0$, we observe the same hierarchy of correlation strengths for $s - s' = 1$ under *No reset*, *MLR*, and *DQLR*. *No reset* performs the worst with correlation strengths at about 1% for all correlation distances studied here. *MLR* performs considerably better, but still has measurable correlations at time distances as large as 6 cycles. Given that we desire zero correlation at all time distances, *DQLR* most closely approximates this condition, with $|\bar{p}_{t,t'}|$ at 0.1% for distance-2 correlations, and otherwise well below 0.1% for longer time distances.

These observed correlation strengths suggest that if scalable QEC requires near-independent errors, complete leakage removal on all qubits must be carried out in some

form. The *DQLR* strategy presented in this work provides one pathway to reaching that requirement. As a corollary, we show that *No reset* and *MLR* cannot support this requirement under existing leakage rates and transport mechanisms.

-
- [1] Kelly, J. *et al.* State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015). URL <http://www.nature.com/articles/nature14270>. Publisher: Nature Publishing Group.
 - [2] McEwen, M. *et al.* Removing leakage-induced correlated errors in superconducting quantum error correction. *Nature Communications* **12**, 1761 (2021). URL <http://www.nature.com/articles/s41467-021-21982-y>.
 - [3] Google Quantum AI *et al.* Suppressing quantum errors by scaling a surface code logical qubit (2022). URL <https://arxiv.org/abs/2207.06431>.
 - [4] O’Brien, T. E., Tarasinski, B. & DiCarlo, L. Density-matrix simulation of small surface codes under current and projected experimental noise. *npj Quantum Information* **3**, 39 (2017). URL <http://www.nature.com/articles/s41534-017-0039-x>. Publisher: Nature Publishing Group.
 - [5] Google Quantum AI *et al.* Exponential suppression of bit or phase errors with cyclic error correction. *Nature* **595**, 383–387 (2021). URL <http://www.nature.com/articles/s41586-021-03588-y>. eprint: 2102.06132.
 - [6] Fowler, A. G. Coping with qubit leakage in topological codes. *Physical Review A* **88**, 042308 (2013). URL <https://link.aps.org/doi/10.1103/PhysRevA.88.042308>. Publisher: American Physical Society.