nature portfolio

Peer Review File

An entangling gate for dual-rail erasure qubits

Corresponding Author: Dr Kevin Chou

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
The paper by Quantum Circuits, Inc. Team experimentally demonstrated the implementation of CZ entangling gate between two dual-rail encoded qubits supported in four superconducting cavities. The dominant error, leakage, can be detected and converted to erasure through the end-of-the-line measurements, with an estimated 0.5% error rate per gate. The remaining error has a Pauli bias with infidelity about 0.1%. With this, the erasure-dominant error structure is preserved in the gate, which will be useful in further quantum error correction (QEC) steps.

To our knowledge, this work is the first realization of the entangling gates with the erasure-dominant error structure in the superconducting platform. The overall CZ gate fidelity is higher than that demonstrated in Rydberg platform with a similar erasure conversion feature [27]. The study is comprehensive, with a detailed discussion on the topics including pulse calibration, fidelity estimation through randomized benchmarking, error structure analysis and simulation, etc. The pulse sequences in the SWS procedure are well-designed to avoid the effect of unwanted noise. In the writing aspect, we believe it is clear enough and easy to follow.

Despite the above merits, we notice that a full demonstration of mid-circuit-level leakage detection and erasure conversion is still missing in this work. We remark that, in the former Rydberg platform experiment [27], they have demonstrated the mid-circuit leakage detection. We believe such a missing gadget is very important in the QEC contexts, as we will elaborate below.

a). The mid-circuit level erasure conversion helps to locate the erroneous qubit in both space and time. To our knowledge, every threshold-improvement claim based on erasure conversion is made with the assumption that leakage can be detected and converted immediately after every entangling gate (see [10], [11], and a recent one [arXiv:2502.20558]). There lacks the evidence if the QEC threshold can still get improved with the leakage error only detected at the end of the whole circuit.

b). The missing of mid-circuit detection also questions the claim in the paper that “leakage propagation is particularly benign”. In this work, CZ only acts on two specific dual-rail qubits (since here are only two), and by no means the error will propagate. If we think in the QEC context where entangling gates can be performed between one qubit and several others, then a leakage in a qubit without immediate detection will make all following gates involving that qubit invalid. Therefore, errors will quickly propagate to all other qubits that are supposed to perform entangling gates with the leaked one. For comparison, Pauli errors may not propagate as fast as this, like Z error commutes with CZ gate so it will not propagate to another qubit.

c). We also have doubts on the feasibility to perform mid-circuit detection in the current platform. From the two lower plots in Fig 2(c) and 2(d), it seems that the y-axis intercepts of the two curves are only 0.9 and 0.6 for the postselected fraction. We expect the authors to explain the large deviation to 100% fraction in the zero CZ/Clifford gate limit. It makes us to wonder if the readout fidelity is too low to support repetitive measurements, since we believe that the least 10%-40% discard rate per measurement is not scalable.

On the other hand, the authors may present a systematic comparison and resource estimation to show the benefits of using the current erasure-dominant superconducting qubits: like the comparison of threshold or qubit resource overhead with

suitable QEC codes? We remark that, there are higher reported CZ gate fidelities in other superconducting platforms. With fluxonium, people have reported the CZ infidelity below 0.1% (see e.g., Ding et al, PRX 13, 031035 (2023)). This is even smaller than the fidelity here after postselection. Though the authors may argue that noise is structured here, we still have concern if (0.5% erasure + 0.1% Pauli Z bias) can beat the only 0.1% error provided by other devices in any QEC simulation. We expect the authors to address the tradeoff between the more structured errors and the overall fidelities here.

Here are also some minor comments regarding the phrase or technical details:

1. The word “bias-preserving” in the title is somewhat misleading. We understand from the text that the “bias” here means “erasure-to-Pauli bias” or erasure-dominant error structure, but we think usually “bias-preserving” refers to the bias between different Pauli operators (like Z to X) (see, Guillaud and Mirrahimi, PRX 9, 041053 (2019)). We suggest rephrasing the title to eliminate possible misunderstanding.

2. The abstract mentioned flag-qubit, but there is no technical discussion on it in the text. Since flag-qubit can be included in the QEC framework, we suggest to directly remove it, or add sufficient discussion in the text.

3. The leakage channel in Eq. (E1), and the leakage jump operator L_c, L_t above Eq. (F7) are wrong. Each term in the Kraus operator K^x should not be coherently summed. Otherwise, the pure state (|0_L> - |1_L>)/sqrt(2) will never decay. The similar argument also applies on L_c, L_t.

4. The notation below Eq. (F4) is confusing. Why ZZ(\phi) is a product of two single-qubit gates instead of an entangling gate itself? And what is \theta here?

Referee #2

(Remarks to the Author)
The authors present a novel error-detected entangling gate between two superconducting dual-rail qubits, achieving groundbreaking performance. While the bare gate infidelity is on the order of 0.5%, the vast majority of these errors are photon loss errors, which can be detected. This "erasure" capability is highly beneficial in an error-correction setting, as error thresholds and effective code distances are increased. Crucially, the gate errors conditioned on no erasures are much lower (~0.1%). Additionally, the remaining errors possess a unique structure that will enhance performance in error correction settings, such as a strong noise bias and significant asymmetry between control and target errors.

Realizing high-fidelity entangling gates with a desirable noise structure compatible with erasure QEC is a very complex challenge - one successfully tackled in this work, which constitutes a big step forward from the authors' work from last year on idling dual-rail qubits (Chou et al., Nat Phys 2024). An important insight is that a modification of the operation demonstrated in Ref. 40 provides the necessary building block for dual-rail entangling gates. The resulting gate performance is impressive, with unprecedented numbers like 1e-6 bit-flip errors and 1e-4 phase-flip errors. The authors compellingly argue that this gate will provide an error correction gain of approximately x10 with increased code distance - significantly higher than the x2 factor recently achieved with superconducting transmon qubits and neutral atoms. Therefore, I would consider this paper a milestone result.

The paper itself is well-written, the methods are solid, and the data are convincing. As detailed above, the insights leading to this work are profound, and the results make an compelling case for scaling this system into a full surface code to demonstrate the predicted QEC gains. The demonstrated work already exceeds all previously published state-of-the art results on this particular platform, to the best of my knowledge. One challenge of this platform, however, is that the physical components and their integration may not easily lend themselves to scaling to the required tens of dual-rail qubits, let alone to hundreds. Nonetheless, the fundamental scientific insights stand on their own, and the methods and principles developed here could be immediately adapted to a more scalable superconducting qubit framework, where similar performance might be expected.
I anticipate that the insights into error structure and error propagation gained from this work will significantly benefit not only superconducting qubits, but also other quantum computing platforms. For these reasons, I recommend publication in Nature nearly as-is.

Specific comments and questions:

1) The references to figure 3 in the text seem incorrect, e.g., 3b instead of 3a.

2) The authors write, "We note that although the CZ gate preserves the Pauli noise-bias, we do not anticipate using a full gate set that preserves Pauli noise-bias". This statement is unclear. Since a universal gate set is not required for QEC, do the authors mean a limited bias-preserving gate set required for QEC is unavailable (e.g., H is never bias preserving)? Clarification or a citation to a relevant paper would help. For example, Guillaud et al," Physical Review X 9.4, 041053 (2019) prove a similar statement for two-level systems in their appendix.

3) The authors should mention why they chose to prepare the target with sqrt(X) in 3c, and clarify why this procedure does not include an echo X pulse in the middle of the sequence, unlike other experiments.

4) In the text referring to 4f, Appendix F should be cited to explain the off-diagonal elements. It might be better not to refer to

this channel as "closely matching a pure dephasing channel", as negligible off-diagonal terms would then be expected.

5) I found no discussion of the prevalence of thermal photons in the coupler or their effect on gate performance (for example, photon number would not be conserved). Appendix H seems to address failures of swapping an excitation back, which is distinct from pure heating.

6) In appendix A, the authors should explain the dominant mechanisms responsible for the significant SPAM error of 6-7%.

7) The authors use end-of-line erasure detection, but mid-circuit erasure measurements are required when implementing QEC, as noted by the authors. What is the dephasing error probability when applying erasure detection, particularly to the target qubit? Likewise, the always-on cross-Kerr of 6.6 kHz might cause error accumulation during stabilizer measurements. Are these effects accounted for in the estimate of the error-correction gain?

8) There is a typo in Ref. [41]

Referee #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4

(Remarks to the Author)
The manuscript demonstrates a high-fidelity CZ gate between two superconducting dual-rail qubits (control-target) using the SWAP-wait-SWAP protocol. After excluding the shots where erasure error events happened, a CZ gate fidelity near 99.9% was achieved. This is the first demonstration of a highfidelity CZ gate between two erasure qubits, giving credence to its viability as an architecture. The results are based on solid theory and careful engineering.

Although the result shown in this paper is an important milestone for dual-rail qubits, we believe this manuscript does not meet the stringent bar for Nature. There have been several two-qubit-gate demonstrations with superconducting qubits over 99.9% fidelity, with considerably faster gate times and without the need for postselection. Moreover, we believe a two-qubit gate demonstration alone is not groundbreaking enough to merit publication in Nature, unless it shows a significant advance in fidelity, which is not the case for this article. This is not meant to detract from the quality of this work, but rather is a sign of the maturity of the field. We feel this paper would be more appropriate for a more specialized journal.

In addition, we have comments regarding some of the main results from the paper:
• In FIG. 2d., the authors use a linear fit to the RB curves for shorter depth instead of applying an exponential fit to the entire result because there is a complication in the fitting. However, the postselected fraction curve appears to visually match an exponential decay. The author should provide more justification for not trying an exponential fit to the postselected data.
• In FIG. 6e., from what we understand, each SWS must contribute a conditional phase of 180 degrees. But in the figure, an even number of SWS results in a -180-degree phase, which was confusing. We think further explanation of this figure would help the reader's understanding.
• The 0.1% CZ gate error derived from interleaved RB lacks uncertainty quantification. Additionally, it also lacks the fidelities of the respective reference and interleaved sequences, as well as their uncertainties. We believe it is important to include these numbers to justify the 0.1% CZ gate error. • In AppA, FIG. 5., the authors should clarify how the single-qubit-gate fidelity has been obtained from the randomized benchmarking result. It is unclear whether the fidelity was calculated from the data before or after postselection.
• In AppA, FIG. 5., the postselected fraction of the control qubit data implies simultaneous RB has a better fidelity because the red square markers are consistently higher than the red circles. However, it is reported as the simultaneous RB sequence fidelity is higher than the individual RB sequence fidelity. Next, we offer the following suggestions to improve the manuscript:
• The authors could try leakage randomized benchmarking analysis for their randomized benchmarking protocol to better estimate the effect of leakage error to the CZ gate fidelity.
• The authors could try increasing the number of Cliffords in their single-qubit randomized benchmarking to see the sequence saturate. For instance, for a high-fidelity single-qubit gate near 99.99%, it is not uncommon to benchmark using up to 10,000 Cliffords.
• The authors could try increasing the number of randomized sequences to improve the randomization.
• Showing the stability of the two-qubit gate over extended periods of time would demonstrate the robustness of the implemented gate.

Finally, we have some minor suggestions to improve clarity or fix suspected typos:
• Citations: In the introduction, we suggest citing 2 of the seminal proposals for dual-rail qubits: o Chuang and Yamamoto, Phys. Rev. A 52, 3489 (1995) o Chuang and Yamamoto, Phys. Rev. Lett. 76, 4281 (1996)
• Page 22: "we also know the value of t and $\phi = \chi t$" – from context should this be ϕ instead of ϕ? Or are the angles in Fig. 11 mislabeled?
• On page 4, there seem to be two different numbers, 0.53% and 0.69%, that describe the total erasure probability per CZ gate. Which one is correct? Quantification of uncertainties would aid in comparing the numbers extracted from the two methods.
• Page 7: "high postselected fidelity (99.97% - 99.90%)" - should the first value be 99.79%?

• AppC: The authors mention a CZ gate on-off ratio of 230. However, the importance of CZ gate on-off ratio in general is unclear or how this value compares to other implementations. For instance, in architectures where idle ZZ can be tuned to 0, this ratio is effectively infinite within experimental resolution.
• AppF, 2: Leakage during CZ gate. Some definitions and explanations in this section are confusing. We were unclear about the following:
o First, the definition of CZ and CZ(ϕ) is unclear. Does CZ(ϕ) correspond to a CPhase gate? It feels nonstandard to put the name “CZ” when the conditional phase is not 180 degrees.
o In Eq. F1, CZ seems to be defined as CZ = $|11\rangle\langle 11|$. In the paragraph below, CZ is redefined as CZ = CZ(π)= ($I + Z + Z + ZZ$). These two definitions do not seem self-consistent and correctly describe the CZ-gate interaction.
o Finally, the definition of θ and its relationship with ϕ is unclear; θ appeared in Eq. F4 without explanation. Is it a Bloch angle? Is it the angle defined in AppK?

Referee #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
In the previous round of review, we mainly expressed our concern on the missing of mid-circuit erasure check, the high SPAM error rate and the related scalability issues, and the QEC performance comparison between the structured noise (erasure dominant + Pauli Z bias) and a high-fidelity generic noise model. We appreciate the authors' enormous efforts on the clarification of the major focus of this work (engineering and benchmarking this novel 2-qubit gate with erasure-dominant error structure), the explanation for the origins of the SPAM error, as well as a detailed simulation on the QEC performance under the structured noise model.

We think the results of this work is fairly strengthened after this revision and we are convinced that the novel error structure engineered from the dual-rail CZ entangling gate could be beneficial for the QEC performance compared with the general depolarized noise model, provided that the SPAM and mid-circuit measurement fidelities are further improved.

On the other hand, despite that the authors argue that there is no need to perform mid-circuit erasure check after every CZ gate in the QEC setup, it is still a necessary gadget after every stabilizer measurement round, and therefore the characterization of its performance is also important in order to make QEC advantage claim for the dual-rail framework. We checked one related paper mentioned by the authors [S.de Graaf, S. Xue et al, npj QI 11(1) (2025)] and found out that the current infidelities for the erasure check (~3-4% erasure or false-negative) is still a little higher than what is used in the QEC simulation. The 0.3% Pauli error from the check is also not included in the simulation. Therefore, we think there is still sufficient evidence to believe that, at the current stage of the hardware performance, it is hard to make any claim that the dual-rail framework could provide QEC advantages compared with other architectures based on superconducting qubits.

In the meantime, while we agree that "for the purpose of this work" (benchmarking CZ gate) the high SPAM error is acceptable, we still keep our concern that it will become a problem when people want to scale up the system in the future. It will be good to see if this can be reduced to 1%-level while maintaining the gate performance for future hardware.

Referee #2

(Remarks to the Author)
recommends publication

Referee #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4

(Remarks to the Author)
Summary:
The authors have clearly addressed most of the questions raised by us and the other reviewers. In addition, the revised manuscript includes two significant claims.

• Mid-circuit erasure checks are not necessarily vital during the error correction cycle in the proposed dual-rail architecture.
• The dual-rail architecture and the CZ gate presented in the paper can, in principle, improve distance scaling compared to a standard two-qubit gate of similar fidelity.

These claims were substantiated through numerical simulations in the manuscript, and we also found a corroborating paper [PRX Quantum 6, 040354 (2025)] that is relevant to the first point. The importance of these findings extends beyond the creation of novel, high-fidelity two-qubit gates and deserves attention from the broader community. Whether it belongs in Nature or a more specialized, but still very high-impact, journal like Nature Physics is an editorial decision.

Minor technical points:
The manuscript needs more explanation regarding the large false-positive error rate (6% according to the Appendix). False-positive errors have a nonnegligible effect on the error threshold under erasure checks, and 6% may not be tolerable for error correction, as relevant experiments [PRX 14, 011051 (2024)] or numerical simulation [PRX Quantum 6, 040354 (2025)] target an error of 1% or less, including the current manuscript. From the response, the authors seem to be aware of this issue, but we believe paper should clarify the conditions that need to be satisfied to achieve the targeted 1% false-positive error rate.

In Figure 2c, which illustrates Bell-state tomography, the authors speculate that the quadratic decrease in fidelity and purity at larger gate depth may result from drifts in calibration parameters or fluctuation in frequency. However, it is not clear that parameter drift would be sufficiently pronounced over the timescale of these specific measurements to cause visible effects. Could there be other plausible explanations such as, for instance, the accumulation of erasure errors during the CZ gate that results in dephasing or an increase in undetected erasure error events?

There is a small typo in Appendix N: $|0\rangle|1\rangle$ should be "$|01\rangle$ state" for consistency. Also, in the below sentence, “In contrast, the evolution of the $|01\rangle$ dual rail” should be $|00\rangle$ instead of $|01\rangle$.

On page 5, “In Appendix K we performed repeated IRB” should be “In Appendix L.”

Referee #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee responses and summary of revisions
Manuscript number: 2025-05-11044

(Dated: February 17, 2026)

OVERVIEW OF RESPONSES

We thank all of the reviewers for their insightful and helpful comments on our manuscript. We note that all reviewers have highlighted the high quality of our work and its significance for dual-rail erasure qubits.

In this document, we directly respond to the main critique of our work and provide evidence that the two-qubit gate construction and dual-rail qubit approach provides immediate and significant benefits for quantum error correction (QEC). The full explanation as to how the noise properties of our dual-rail cavity qubits improves QEC, and the associated simulations, is the subject of an upcoming publication. Nevertheless, to provide suitable context for the reviewers and readers, we have added a section to this response letter and two new Appendix sections (Appendix P and Q) that demonstrate how the error structure of the SWS gate gives rise to improvements in a simulated dual-rail surface code. We believe these results will alleviate the concerns from the editor and reviewers. We also address the numerous other comments made by the reviewers, making changes to the manuscript where appropriate. In summary, we believe our work represents a significant contribution towards practical realization of an error-corrected quantum computer. Furthermore, this work provides strong experimental validation of a crucial component of the dual-rail cavity architecture. As such we argue, and sincerely hope the Editors and reviewers agree, that this work now qualifies for publication in the journal *Nature*.

REVIEWER A

Reviewer Comment A.1 — The paper by Quantum Circuits, Inc. Team experimentally demonstrated the implementation of CZ entangling gate between two dual-rail encoded qubits supported in four superconducting cavities. The dominant error, leakage, can be detected and converted to erasure through the end-of-the-line measurements, with an estimated 0.5% error rate per gate. The remaining error has a Pauli bias with infidelity about 0.1%. With this, the erasure-dominant error structure is preserved in the gate, which will be useful in further quantum error correction (QEC) steps.

To our knowledge, this work is the first realization of the entangling gates with the erasure-dominant error structure in the superconducting platform. The overall CZ gate fidelity is higher than that demonstrated in Rydberg platform with a similar erasure conversion feature [27]. The study is comprehensive, with a detailed discussion on the topics including pulse calibration, fidelity estimation through randomized benchmarking, error structure analysis and simulation, etc. The pulse sequences in the SWS procedure are well-designed to avoid the effect of unwanted noise. In the writing aspect, we believe it is clear enough and easy to follow.

Reply: We agree with the reviewer's assessment. We would also like to comment that the careful

error metrology presented in this work for this specific two-qubit gate is increasingly necessary, particularly in the context of quantum error correction. Comprehensive error analysis is essential for accurate noise models and accurate modeling for large-scale quantum systems.

Reviewer Comment A.2 — Despite the above merits, we notice that a full demonstration of mid-circuit-level leakage detection and erasure conversion is still missing in this work. We remark that, in the former Rydberg platform experiment [27], they have demonstrated the mid-circuit leakage detection. We believe such a missing gadget is very important in the QEC contexts, as we will elaborate below.

Reply: While we do agree that leakage detection and erasure conversion is necessary for full error correction, we would like to highlight that the focus of this result is a detailed characterization and benchmarking of a novel two-qubit gate. The integration of gate operations and mid-circuit leakage detection measurements is the focus of ongoing work and we argue why these gadgets are not necessary to characterize and benchmark our two-qubit gate.

For our platform we are able to utilize leakage detection capabilities during the end-of-line measurement to efficiently detect when a decay event has occurred. Additionally, as shown in K. Chou et al, Nature Physics (2024), our dual-rail cavity qubits have been measured to have ultra low-rates of photon gain (equivalently seepage back into the codespace). This means that once a decay error has happened, then the qubit will remain in this erased state until the end of the sequence when it will be detected by the end-of-line measurement. Therefore, in this work we are able to both effectively detect erasures during the experiment and reliably characterize the CZ gate with a single erasure detection measurement at the end of the experiment.

We can also argue that mid-circuit leakage detection operations are less important than previously thought within the context of a dual-rail surface code. This is because they may be delayed to the end of each stabilizer round (occurring after every four CZ gates instead of after every CZ gate) without much loss of threshold or effective code distance (see our simulations in the new Appendix Q and Section I A of this document). Moreover, the need of this operation may be obviated entirely in a quantum error correction context, either by a successive CNOT and anti-CNOT gate as proposed in [arXiv:2509.13247] (2025). Or by teleporting the state of the qubits back and forth between data and ancilla qubits at the end of each stabilizer round (and detecting erasures instead via end-of-the-line measurements, which may be destructive.)

Reviewer Comment A.3 — a). The mid-circuit level erasure conversion helps to locate the erroneous qubit in both space and time. To our knowledge, every threshold-improvement claim based on erasure conversion is made with the assumption that leakage can be detected and converted immediately after every entangling gate (see [10], [11], and a recent one [arXiv:2502.20558]). There lacks the evidence if the QEC threshold can still get improved with the leakage error only detected at the end of the whole circuit.

Reply: We counter this point by providing new, strong evidence that the QEC threshold remains high for erasures, when leakage errors are only detected at the end of stabilizer measurement rounds for our particular leakage error model. The property that enables these 'delayed' erasure checks is the fact that leakage errors to the vacuum state 'turn-off' the gate (and all subsequent gates until the leaked qubit is reset). We incorporate this property into Stim surface code simulations, and present the results in Section I A and I B of this document and in Appendix Q of the manuscript. We compare perfect erasure checks after every CZ gate vs. delayed erasure

checks and extract similar surface code error thresholds and error rate scaling for both. In reality, since these erasure checks are also error-prone, we expect better surface code performance with delayed erasure checks compared to the status quo where erasure checks are performed after every gate.

Reviewer Comment A.4 — b). The missing of mid-circuit detection also questions the claim in the paper that "leakage propagation is particularly benign". In this work, CZ only acts on two specific dual-rail qubits (since here are only two), and by no means the error will propagate. If we think in the QEC context where entangling gates can be performed between one qubit and several others, then a leakage in a qubit without immediate detection will make all following gates involving that qubit invalid. Therefore, errors will quickly propagate to all other qubits that are supposed to perform entangling gates with the leaked one. For comparison, Pauli errors may not propagate as fast as this, like Z error commutes with CZ gate so it will not propagate to another qubit.

Reply: We clarify here what we mean by "leakage propagation is particularly benign". By this, we specifically mean that leakage to the vacuum state causes subsequent CZ gates involving the leaked qubit to be replaced with identity operations. This is subtly different to a common leakage propagation model, where CZ gates involving a leaked qubit result in depolarizing or dephasing errors on the qubit that is not leaked. We show in Section I A and in Appendix Q of the manuscript that while delaying leakage detection does cause leakage to propagate for up to four timesteps (in the surface code), the error syndromes under our model remain rather simple and easy to decode, which is illustrated by the error scaling and high threshold which is $\sim 5\%$ per CZ gate.

Reviewer Comment A.5 — c). We also have doubts on the feasibility to perform mid-circuit detection in the current platform. From the two lower plots in Fig 2(c) and 2(d), it seems that the y-axis intercepts of the two curves are only 0.9 and 0.6 for the postselected fraction. We expect the authors to explain the large deviation to 100% fraction in the zero CZ/Clifford gate limit. It makes us to wonder if the readout fidelity is too low to support repetitive measurements, since we believe that the least 10%-40% discard rate per measurement is not scalable.

Reply: This work focused on gate calibration and error metrology and we optimized the system for low logical misassignment and leakage detection to quantify very small gate errors. This choice comes at the cost of higher erasure errors for our SPAM. We argue that, for the purpose of this work, this is an acceptable and well understood trade-off. The zero-gate postselection fraction can be understood by the SPAM erasure assignments (see Table III in the Appendix). For state tomography we employ a 1-round measurement and in agreement with our independent measurements extract a total two-qubit postselection fraction of 88%. For RB, we employ a 2-round measurement to improve our leakage detection efficiency which are expected to be much higher due to the circuit depths used for RB. We provided more information about SPAM for dual-rail cavity qubits in our previous work, K. Chou et al Nature Physics (2024).

Looking ahead, we believe the optimization for faster dual-rail measurements to be relatively straightforward–for example, by speeding up the transmon readout measurement, which currently is slow, taking $2-3$ μs (State-of-the-art is typically ~ 500ns). We expect to target erasure assignments to approach the 1%-level for future hardware.

Furthermore, the measurement settings for an actual surface code ancilla readout mid-circuit readout would be much different, and one would trade off higher measurement error for shorter duration and lower induced erasures. See Section I B for what 'discard rates' or ancilla readout false positives are tolerable in a surface code and how they affect QEC performance.

Reviewer Comment A.6 — On the other hand, the authors may present a systematic comparison and resource estimation to show the benefits of using the current erasure-dominant superconducting qubits: like the comparison of threshold or qubit resource overhead with suitable QEC codes? We remark that, there are higher reported CZ gate fidelities in other superconducting platforms. With fluxonium, people have reported the CZ infidelity below 0.1% (see e.g., Ding et al, PRX 13, 031035 (2023)). This is even smaller than the fidelity here after postselection. Though the authors may argue that noise is structured here, we still have concern if (0.5% erasure + 0.1% Pauli Z bias) can beat the only 0.1% error provided by other devices in any QEC simulation. We expect the authors to address the tradeoff between the more structured errors and the overall fidelities here.

Reply: We thank the referee for this important and insightful comment. Whilst we can make qualitative arguments as to why structured noise should lead to QEC benefits, we agree that performing QEC simulations is the best way to provide a direct quantitative comparison. We would like to direct the referee to Section I C 4 where we compare surface code QEC scaling performance between an error model with depolarizing noise after each CZ gate with strength 0.1%, compared with the (0.5% erasure + 0.06% Pauli Z bias) model for our gate. We quantify our scaling performance with the metric Λ, defined to be the factor of reduction in (per-round) Logical error rate in a surface code memory each time the code distance increases by two. For the depolarizing noise model, since the threshold is about 1%, we obtain $\Lambda \approx 14$. However, for the SWS gate error model, we obtain $\Lambda \approx 27$, a value almost twice as large, which quantifies the benefits of structured two-qubit gate noise. These noise models are useful for comparing two-qubit gates against each other but more practical comparisons must also consider other error channels such as faulty measurements, and for us, faulty erasure checks as well. These more representative error models will be detailed in our upcoming publication. To provide additional support for our claims in the manuscript we have added Appendix P and Q.

Reviewer Comment A.7 — 1. The word "bias-preserving" in the title is somewhat misleading. We understand from the text that the "bias" here means "erasure-to-Pauli bias" or erasure-dominant error structure, but we think usually "bias-preserving" refers to the bias between different Pauli operators (like Z to X) (see, Guillaud and Mirrahimi, PRX 9, 041053 (2019)). We suggest rephrasing the title to eliminate possible misunderstanding.

Reply: Our two-qubit gate demonstrates both erasure-to-Pauli bias *and* Pauli-error bias with Z-type errors dominant over X-type errors. Furthermore, we do show that the errors are preserved through application of the gate–that is X or Z errors before the gate remain X or Z errors after the gate, which is reflected in the pronounced asymmetry in the measured bitflip and phaseflip rates. We also note that our gate exhibits the same type of noise bias (CZ preserves dephasing bias, Hadamard gates do not) that was first studied in Aliferis, Preskill, Phys. Rev. A 78, 052331 (2008), and was the first work to show the benefits of Pauli noise bias in a quantum error correction context. As such, we argue that the use of "bias-preserving" in the title is appropriate, and maybe serves to remind readers the scope of Pauli noise bias historically extends beyond

the noise bias commonly associated with cat qubits. Although we lack a full set of (Pauli) bias-preserving gates, we believe that our existing noise bias is still important to achieving impressive QEC performance with dual-rail cavity qubits, as reflected by the thresholds in Fig. 3 and Fig. 4 obtained in our simulation results. As such, we want to highlight this gate property in the title of our manuscript.

Reviewer Comment A.8 — 2. The abstract mentioned flag-qubit, but there is no technical discussion on it in the text. Since flag-qubit can be included in the QEC framework, we suggest to directly remove it, or add sufficient discussion in the text.

Reply: We have decided to remove this from the text. There will be separate, upcoming work using this gate in the context of flag qubits, but it will be too technical to explain the necessary details in this work.

Reviewer Comment A.9 — 3. The leakage channel in Eq. (E1), and the leakage jump operator L_c, L_t above Eq. (F7) are wrong. Each term in the Kraus operator K^x should not be coherently summed. Otherwise, the pure state $\left(\left(|0_L\rangle - |1_L\rangle\right)/\sqrt{2}\right)$ will never decay. The similar argument also applies on L_c, L_t.

Reply: We appreciate finding this issue and have made corrections in the supplement such that the correct set of Kraus operators are used to describe this error channel.

Reviewer Comment A.10 — 4. The notation below Eq. (F4) is confusing. Why $ZZ(\phi)$ is a product of two single-qubit gates instead of an entangling gate itself? And what is θ here?

Reply: We appreciate finding this issue and have addressed in the text to define gate angles more clearly and remove any ambiguity about the unitary described by $ZZ(\phi)$. In general, any Pauli operator $\hat{O}$, written as $\hat{O}(\theta)$ is shorthand for $\cos(\theta/2)\mathbb{1} - i\sin(\theta/2)\hat{O}$. We also rewrote section F2 of the appendix to make the derivation of the error channel easier to follow. (we also noticed a small mistake in one of the coefficients of the error channel which we have corrected.)

REVIEWER B

Reviewer Comment B.1 — The authors present a novel error-detected entangling gate between two superconducting dual-rail qubits, achieving groundbreaking performance. While the bare gate infidelity is on the order of 0.5%, the vast majority of these errors are photon loss errors, which can be detected. This "erasure" capability is highly beneficial in an error-correction setting, as error thresholds and effective code distances are increased. Crucially, the gate errors conditioned on no erasures are much lower (0.1%). Additionally, the remaining errors possess a unique structure that will enhance performance in error correction settings, such as a strong noise bias and significant asymmetry between control and target errors.

Realizing high-fidelity entangling gates with a desirable noise structure compatible with erasure QEC is a very complex challenge - one successfully tackled in this work, which constitutes a big step forward from the authors' work from last year on idling dual-rail qubits

(Chou et al., Nat Phys 2024). An important insight is that a modification of the operation demonstrated in Ref. 40 provides the necessary building block for dual-rail entangling gates. The resulting gate performance is impressive, with unprecedented numbers like 1e-6 bit-flip errors and 1e-4 phase-flip errors. The authors compellingly argue that this gate will provide an error correction gain of approximately x10 with increased code distance - significantly higher than the x2 factor recently achieved with superconducting transmon qubits and neutral atoms. Therefore, I would consider this paper a milestone result.

Reply: We appreciate the Reviewer's close examination of the manuscript and their positive assessment of its impact.

Reviewer Comment B.2 — The paper itself is well-written, the methods are solid, and the data are convincing. As detailed above, the insights leading to this work are profound, and the results make an compelling case for scaling this system into a full surface code to demonstrate the predicted QEC gains. The demonstrated work already exceeds all previously published state-of-the art results on this particular platform, to the best of my knowledge. One challenge of this platform, however, is that the physical components and their integration may not easily lend themselves to scaling to the required tens of dual-rail qubits, let alone to hundreds. Nonetheless, the fundamental scientific insights stand on their own, and the methods and principles developed here could be immediately adapted to a more scalable superconducting qubit framework, where similar performance might be expected.

I anticipate that the insights into error structure and error propagation gained from this work will significantly benefit not only superconducting qubits, but also other quantum computing platforms. For these reasons, I recommend publication in Nature nearly as-is.

Reply: We appreciate the Reviewers' careful reading of our work and do concur with the assessment of its significance.

With regards to scaling to tens of hundreds of dual-rail qubits for error correction, we believe that the challenges we face are comparable to those faced in more popular superconducting qubit platforms. When we tally up the total number of control lines and transmon-like elements (e.g. our SQUID-based couplers) that would be required for a surface code plaquette, this is is usually similar to the numbers required for a transmon-based surface code architecture based on tunable couplers. Moreover, in this work, we opt for specific hardware complexities (e.g. a readout line for every single cavity) that make it easier for detailed characterization of the two-qubit gate, but can be simplified when scaling up an actual surface code.

Reviewer Comment B.3 — 1) The references to figure 3 in the text seem incorrect, e.g., 3b instead of 3a.

Reply: Thank you for pointing out this mistake, we have made the correction.

Reviewer Comment B.4 — 2) The authors write, "We note that although the CZ gate preserves the Pauli noise-bias, we do not anticipate using a full gate set that preserves Pauli noise-bias". This statement is unclear. Since a universal gate set is not required for QEC, do the authors mean a limited bias-preserving gate set required for QEC is unavailable (e.g., H is never bias preserving)? Clarification or a citation to a relevant paper would help. For example, Guillaud et al," Physical Review X 9.4, 041053 (2019) prove a similar statement for two-level systems in their appendix.

Reply: This statement was intended place our work in the broader context of bias-preserving gates, including the Guillaud et al. work you reference. The main claim for our work is that the SWS CZ gate does exhibit both a erasure and Pauli noise bias, but importantly does not imply that we have demonstrated a fully bias-preserving *gate set*. Indeed, it is generally considered that the full gate set for dual-rail qubits contains single qubit gates (such as Hadamards) that *do not* preserve the Pauli noise bias.

The reviewer identifies the correct reason why a fully bias-preserving gate-set is currently unavailable for our dual-rail qubits. Instead we are referring to bias-preserving Pauli noise as studied in Aliferis, Preskill, Phys. Rev. A 78, 052331 (2008), who were among the first to introduce the concept of Pauli noise bias for QEC. We find that in the surface code, having a dephasing noise bias after our CZ gates approximately doubles the surface code threshold to Pauli errors (See Section I C 2). Similar improvements have been reported under dephasing-dominated CZ gate error models, despite lacking a full bias-preserving gate set. (See Darmawan et al., PRX Quantum 2, 030345 (2021) where they obtained a threshold $\sim 1.7\%$)

We have made edits to provide more clarity as to what we mean by a bias-preserving two qubit gate, and have also made some minor changes to the introduction adding the Aliferis, Preskill citation.

Reviewer Comment B.5 — 3) The authors should mention why they chose to prepare the target with sqrt(X) in 3c, and clarify why this procedure does not include an echo X pulse in the middle of the sequence, unlike other experiments.

Reply: This is a good question and we appreciate the opportunity to expand on the experimental design choices for this measurement. The goal is to quantify the bit-flip error rate per CZ gate on the control and target qubits. To isolate this particular error mechanism it is beneficial to modify the circuit used for state tomography. First, we initialize the target (control) qubit in a Z-basis state so that we are not sensitive to any dephasing noise and can directly quantify the fraction of shots where we observe a bit-flip outcome. Second, we opted not to add in echos to eliminate an additional source of bit-flip errors as dephasing during this echo pulse can add extra bit-flip errors. Finally we elected to perform a Ramsey-style sequence on the target qubit to initialize that qubit as a superposition for a more representative quantification of the target (control) -qubit bit-flip error rate. We have implemented some minor edits to clarify the experiment design choices.

Reviewer Comment B.6 — 4) In the text referring to 4f, Appendix F should be cited to explain the off-diagonal elements. It might be better not to refer to this channel as "closely matching a pure dephasing channel", as negligible off-diagonal terms would then be expected.

Reply: We thank the reviewer for this comment and now refer to Appendix F in the caption of figure 4, where the off-diagonal elements are explained. It is worth noting that in an error correction context, these off-diagonal terms are expected to be annihilated by the projective stabilizer measurements (somewhat similar to twirling the noise), meaning this channel can be accurately modeled as a (conditional) dephasing channel.

Reviewer Comment B.7 — 5) I found no discussion of the prevalence of thermal photons in the coupler or their effect on gate performance (for example, photon number would not

be conserved). Appendix H seems to address failures of swapping an excitation back, which is distinct from pure heating.

Reply: The reviewer is correct to point out that thermal excitations in the coupler present an important error mechanism that is distinct from excitations being left in the coupler. Thermal photons present in the coupler at the start of the experimental shot will result in the entire gate sequence failing with probability $\bar{n}_{\text{thermal}}$. This error can be considered as a type of state preparation error, and does not contribute to the error-per-gate metric. (We also note that these kinds of errors also only contribute to the 'SPAM' offset in commonly used two-qubit benchmarking protocols such as randomized benchmarking.) This error mechanism can also induce a correlated error on both control and target dual-rail qubits due to the dispersive interaction causing frequency errors on both qubits. However, we do point out for our results, since the population does not change with repeated gates, it affects the offset and crucially not a per-gate effect. Accurately measuring the small thermal population of the couplers must be done carefully, but we are able to use a Ramsey measurement simultaneously on both control and target dual-rail qubits to bound the thermal population of the coupler to 0.1%.

We have added a new section, Appendix J, that describes this measurement and quantifies the thermal population in the inter-dual-rail coupler.

Reviewer Comment B.8 — 6) In appendix A, the authors should explain the dominant mechanisms responsible for the significant SPAM error of 6-7%.

Reply: System parameters relevant for SPAM are similar to what was used in K. Chou et al. Nature Physics 20, 1454–1460 (2024) and a detailed error budget is provided in that reference (see Appendix IX) to explain the erasure assignment fraction of 6-7%.

Here we will review the error mechanisms that lead to measurement error. The measurement process involves several individual steps on both cavity subsystems: first, a check measurement that checks if the transmon is in the $|g\rangle$ state; second, a photon-number selective π-pulse that flips the transmon if and only if the cavity is in $|1\rangle$; finally, a measurement of the transmon to extract whether there is exactly 1 photon in the cavity or not. Here are the dominant error mechanisms that contribute to SPAM erasures assignments: cavity decay during the protocol, transmon decay or dephasing, cavity state preparation error, and readout classification error.

We do emphasize that SPAM in this system was not optimized to minimize erasure assignment fraction and the system was configured in "characterization mode" for benchmarking the small errors in the two-qubit gate. As stated in Comment A.5, this erasure assignment can be readily decreased in future works, and we believe can achieve erasure assignment at the 1% level or lower.

We have made a short addition to the SPAM section in Appendix A, referring the reader to our previous work for information about the SPAM error budget.

Reviewer Comment B.9 — 7) The authors use end-of-line erasure detection, but mid-circuit erasure measurements are required when implementing QEC, as noted by the authors. What is the dephasing error probability when applying erasure detection, particularly to the target qubit? Likewise, the always-on cross-Kerr of 6.6 kHz might cause error accumulation during stabilizer measurements. Are these effects accounted for in the estimate of the error-correction gain?

Reply: We appreciate the referees questions related to mid-circuit erasure measurements, but would like to reiterate the focus of this paper is on the two-qubit gate. Integration of these erasure checks to complete the QEC toolbox of operations will be the topic of future work. We would also invite the referees to the following previous works on erasure checks for dual-rail cavity qubits: S.de Graaf, S. Xue et al, npj QI 11(1) (2025) and A. Koottandavida, I. Tsioutsios, et al. PRL 132, 180601 (2024).

As a preview for our future work, we can comment on the good questions posed by the Reviewer. The reviewer is correct to point out that the effects of always-on cross-Kerr between neighboring cavities (between different dual-rail qubits) must be accounted for. The always-on cross-Kerr between the control and target qubit is essentially absorbed into the 2Q gate tuneup procedure, and is fully compensated for by simply varying the wait time between the swap pulses until the total entangling phase accumulation is π. The more damaging effect of always-on cross-Kerr is one that is also present in other superconducting qubit platforms: When performing a CZ gate in a surface code context, the ancilla qubit will also perform CPHASE gates with the surrounding 'spectator' qubits due to the always on cross-Kerr. These unwanted CPHASE gates typically have a small angle (~ 0.02 radians). After using stabilizer measurements to effectively Pauli-twirl this error, we expect cross-Kerr to contribute at $\sim 10^{-4}$ to the per-gate Pauli error rate, which we have included in our simulations as an idling error (not shown here, but more detailed simulations suggest $\Lambda \sim 10$ can still be obtained in the presence of this error channel). The usual strategies employed in other qubit platforms to suppress this spectator error also apply to our dual-rail qubits. For example, we can adopt the echo sequence used in Google's XZZX surface code memory experiments (See Fig. S12 of 'Quantum error correction below the surface code threshold', Nature 2024), which partially cancels out the entangling phase acquired from the unwanted CPHASE gates by adding extra X gates between the CZ gates. In the context of the standard stabilizer measurement sequence for a surface code, half of the unwanted CPHASE gates can be completely compensated for by changing the angle of the SWS gates to be slightly less than π. E.g. if a data qubit and an ancilla perform an unwanted CPHASE gate with (measurable) angle 0.02 radians, then adjust the delay of their CZ gate to have an angle of $(\pi - 0.02)$ radians such that the total entangling phase between these qubits over the course of the entire stabilizer measurement sequence totals to π.

Reviewer Comment B.10 — 8) There is a typo in Ref. [41]

Reply: Thank you for pointing this out, we have corrected this typo.

REVIEWER C

Reviewer Comment C.1 — The manuscript demonstrates a high-fidelity CZ gate between two superconducting dual-rail qubits (control-target) using the SWAP-wait-SWAP protocol. After excluding the shots where erasure error events happened, a CZ gate fidelity near 99.9% was achieved. This is the first demonstration of a high-fidelity CZ gate between two erasure qubits, giving credence to its viability as an architecture. The results are based on solid theory and careful engineering. Although the result shown in this paper is an important milestone for dual-rail qubits, we believe this manuscript does not meet the stringent bar for Nature. There have been several two-qubit-gate demonstrations with superconducting qubits

over 99.9% fidelity, with considerably faster gate times and without the need for postselection. Moreover, we believe a two-qubit gate demonstration alone is not groundbreaking enough to merit publication in Nature, unless it shows a significant advance in fidelity, which is not the case for this article. This is not meant to detract from the quality of this work, but rather is a sign of the maturity of the field. We feel this paper would be more appropriate for a more specialized journal.

Reply: We thank the Reviewer for the helpful feedback and have made several improvements to the manuscript to expand on the utility of this two-qubit gate for QEC. We argue that the goal of this paper is not simply about obtaining a higher fidelity gate, but showing that the CZ gate satisfies the properties of a good two-qubit gate for an erasure qubit, which in turns enables the QEC benefits promised in many recent erasure qubit proposals. In a QEC context, errors that are postselected when benchmarking the gate would instead be detected, converted to erasures, and finally corrected, such that we keep 100% of the data. As such, we expect our SWS can offer more compelling performance than other two-qubit gates with similar overall Pauli error rates, due to the error-structure of the gate. We provide strong evidence for this through surface code memory simulations in Section I C as well as Appendix P and Q in the manuscript (also please see our response to A.6). We hope that by showing how this two-qubit gate directly benefits QEC, the Reviewer will agree with the high impact of our work and its suitability to publish in Nature.

Reviewer Comment C.2 — In FIG. 2d., the authors use a linear fit to the RB curves for shorter depth instead of applying an exponential fit to the entire result because there is a complication in the fitting. However, the postselected fraction curve appears to visually match an exponential decay. The author should provide more justification for not trying an exponential fit to the postselected data.

Reply: We do expect–and measure (as the referee observes) an exponential decay behavior for the postselected fraction. Our analysis fits the full exponential decay curve to extract the erasure per CZ gate. On the other hand, the decay curve for the sequence fidelity does not fully saturate at long RB sequences. While we can fit an exponential curve to that data, the fitted saturation value approaches zero instead of 0.25 as expected for fully depolarized, but not erased, states.

We instead elect to fit the short-depth sequence fidelity to a linearized exponential over the first 50 Cliffords as described in Appendix D. In any RB experiment, SPAM and leakage determine the exponential amplitude and offset. Knowing these values in advance is required for our linearized model. We have previously shown high-fidelity measurements in Chou et. al., Nature Physics (2024) that distinguish leaked states from the codespace with errors below 1%. We do not believe that an offset below 0.25, as the trend in Fig. 2d indicates, suggests undetected leakage. Rather, we believe that further investigation is warranted at long sequence depths where erased shots dominate.

Fitting at short depth where counting statistics are not decimated by erasures allows us to extract a fidelity estimate. We describe this procedure in Appendix D. As mentioned above, the exponential offset is an input to the linearized model. Our choice of 0.25 is a pessimistic value that leads to a fidelity underestimate. We note that we are not the first to extract gate fidelity using a linearized RB fit: Ma et. al., Nature (2023) and others do as well. See Extended Data Fig. 3 for their justification of short-depth linear fits.

We have updated Fig. 8 in Appendix E to use a linearized fit to the same Clifford depth as we use for the experimental data analysis. Previously, the figure presented gate infidelities found by

fitting simulated sequence fidelities to an exponential. Those results confirmed accurate fidelity estimation, but may have artificially benefited from the lack of shot noise and ideal long sequence behavior in the density matrix-based simulation. The new version of Fig. 8 hews as close as possible to the analysis performed on the data. It confirms that our linearized fits still produce accurate estimates of gate fidelity. We also added text in Fig. 8 clarifies the fitting procedure.

Reviewer Comment C.3 — In FIG. 6e., from what we understand, each SWS must contribute a conditional phase of 180 degrees. But in the figure, an even number of SWS results in a -180-degree phase, which was confusing. We think further explanation of this figure would help the reader's understanding.

Reply: This is a insightful observation, indeed each SWS gate should impart an entangling phase of 180 degrees. This effect is a result of the π-pulse that occurs in the middle of our calibration sequence. This π-pulse inverts the state of the control qubit and results in the dispersive interaction which enacts the gate to turn off for one half of the sequence. This is the reason we have a 90 degree phase shift on average per CZ gate in the sequence described in Fig. 6e and for even numbers of CZ gates we accrue π shift. We also remark that this protocol of adding the π-pulse in the middle of the sequence enables us to extract only the entangling phase of the CZ gate without needing to calibrate other single qubit phases and also echoes out low frequency noise. (see Gross, J. A. et al. Characterizing coherent errors using matrix-element amplification. npj Quantum Information 10, 123 (2024)). In addition, we have also added a line in Appendix B to make this point clearer.

Reviewer Comment C.4 — The 0.1% CZ gate error derived from interleaved RB lacks uncertainty quantification. Additionally, it also lacks the fidelities of the respective reference and interleaved sequences, as well as their uncertainties. We believe it is important to include these numbers to justify the 0.1% CZ gate error.

Reply: We thank the readers for highlighting this oversight. We have added uncertainties for the overall CZ gate error as extracted by interleaved RB in the main text and in Appendix D and quantify the error per CZ gate to be $0.108(5)\%$. We also have added fit uncertainties to the short-depth (out to $N = 15$) CZ gate error as extracted from Bell state tomography to be $0.029(6)\%$.

Reviewer Comment C.5 — In App A, FIG. 5., the authors should clarify how the single-qubit-gate fidelity has been obtained from the randomized benchmarking result. It is unclear whether the fidelity was calculated from the data before or after postselection.

Reply: For erasure qubits it is important to report two types of data: first the fraction of shots that are assigned as erasures–the "postselected fraction"; and second, the codespace results after postselecting shots where an erasure has been assigned–the "sequence fidelity". Our single-qubit RB data is processed in this fashion–in alignment with all of our other analysis.

Reviewer Comment C.6 — In AppA, FIG. 5., the postselected fraction of the control qubit data implies simultaneous RB has a better fidelity because the red square markers are consistently higher than the red circles. However, it is reported as the simultaneous RB sequence fidelity is higher than the individual RB sequence fidelity.

Reply: Our assessment is that we observe no statistically significant difference in erasure rate when performing simultaneous RB vs. isolated RB. This is shown in the RB data where, for both qubits, the postselected fraction results are equivalent within error bars for the case of individual and simultaneous RB.

The single-qubit gate error is computed after postselecting for erasures. Our results do indeed show that simultaneous RB sequence fidelity is slightly lower than the individual RB sequence fidelity. This result is typically expected due to crosstalk errors inducing additional errors when operating the two qubits simultaneously. Our results do show such crosstalk errors are minimal given small gate error differences at the 10^{-5}-level. For the control qubit we extract $9.0(2) \times 10^{-5}$ (individual) and $1.01(1) \times 10^{-4}$ (simultaneous), respectively. For the target qubit we extract $1.08(1) \times 10^{-4}$ (individual) and $1.36(1) \times 10^{-4}$ (simultaneous), respectively.

Reviewer Comment C.7 — The authors could try leakage randomized benchmarking analysis for their randomized benchmarking protocol to better estimate the effect of leakage error to the CZ gate fidelity.

Reply: Leakage randomized benchmarking (LRB) as described by Wood and Gambetta, PRA (2018) separates the gate fidelity within the codespace from leakage and seepage rates. As shown in our previous work Chou et. al., Nature Physics (2024), the seepage of our DR qubits is at least 1000 times rarer than leakage. This can be seen in Fig. 2d where the probabability of remaining in the codespace converges to zero at high sequence depth. This implies that $A = 0$ in Eq. 9 of Wood and Gambetta, producing a seepage rate approaching zero.

The combination of no appreciable seepage and inert leakage means that shots passing our error detection have only evolved in the codespace. RB in this case only estimates the remaining Pauli errors in the codespace without leakage. This motivates our definition of the postselected gate fidelity in Eq. E5 and obviates the need for LRB. Appendix E contains a thorough discussion of how we incorporate leakage detection into IRB and validate its accuracy.

Reviewer Comment C.8 — The authors could try increasing the number of Cliffords in their single-qubit randomized benchmarking to see the sequence saturate. For instance, for a high-fidelity single-qubit gate near 99.99%, it is not uncommon to benchmark using up to 10,000 Cliffords.

Reply: Whilst the referee makes a good point, measuring out to 10,000 Clifford gates is not practical for us, since the probability of remaining in the codespace is too low to achieve a meaningful result. With no seepage and a measured erasure fraction of 0.1% per single-qubit gate, the un-erased fraction will be vanishingly small at 5×10^{-5} after 10000 gates. Our ability to distinguish leakage states from the codespace is essential in interpreting the linear fit results. Linear fits are not uncommon in RB: Ma et. al., Nature (2023) present single-qubit RB up to only 300 Cliffords and perform linear fits to estimate the fidelity. Our response to comment C.2 includes further discussion of the validity of fitting shorter sequence depths with a linearized exponential.

Reviewer Comment C.9 — The authors could try increasing the number of randomized sequences to improve the randomization.

Reply: The number of random seeds needed to accurately estimate gate fidelity is surprisingly low. Epstein et. al., PRA (2014) show that with as few as 10 random sequences, one can

estimate the true error rate within a factor of two. We confirm via simulations in Appendix E that we achieve much higher accuracy than that with 20 random sequences. Fig. 8 compares the infidelity of a CZ gate computed with Eq. E5 against the value inferred from IRB. The simulation is performed with the exact same sequences as the results presented in Fig. 2d. The results suggest the measured CZ gate infidelity of $0.108(5)\%$ is a slight underestimate of the true infidelity of $0.138(7)\%$.

Reviewer Comment C.10 — Showing the stability of the two-qubit gate over extended periods of time would demonstrate the robustness of the implemented gate.

Reply: We agree that demonstration of longer-term stability will aid in the practical utility of this gate. To this end, we have added Appendix L to address this comment. We perform repeated IRB measurements on multiple two-qubit gate systems that show the gate performance remains stable over many hours.

Reviewer Comment C.11 — Citations: In the introduction, we suggest citing 2 of the seminal proposals for dual-rail qubits: Chuang and Yamamoto, Phys. Rev. A 52, 3489 (1995), Chuang and Yamamoto, Phys. Rev. Lett. 76, 4281 (1996)

Reply: We thank the reviewers for the good suggestion. Both have been added to the manuscript.

Reviewer Comment C.12 — Page 22: "we also know the value of t_{wait} and $\phi_2 = \chi_{bc} t_{\text{wait}}$— from context should this be ϕ_1 instead of ϕ_2? Or are the angles in Fig. 11 mislabeled?

Reply: Thank you for pointing this out. With have corrected the text to read $\phi_1 = \chi_{bc} t_{\text{wait}}$.

Reviewer Comment C.13 — On page 4, there seem to be two different numbers, 0.53% and 0.69%, that describe the total erasure probability per CZ gate. Which one is correct? Quantification of uncertainties would aid in comparing the numbers extracted from the two methods.

Reply: These numbers are extracted from different experiments: 0.53% comes from Bell state tomography while 0.69% is extracted from the full iRB dataset. We have added uncertainties to these quantities.

Reviewer Comment C.14 — Page 7: "high postselected fidelity (99.97% - 99.90%)" - should the first value be 99.79%?

Reply: This is not a typo. In our QST results we extract for short gate depths an error per CZ gate of 0.03%. When measuring out to longer depths with iRB (where we observe nonlinearity in the QST data), we extract a CZ gate error of 0.1%.

Reviewer Comment C.15 — AppC: The authors mention a CZ gate on-off ratio of 230. However, the importance of CZ gate on-off ratio in general is unclear or how this value compares to other implementations. For instance, in architectures where idle ZZ can be tuned to 0, this ratio is effectively infinite within experimental resolution.

Reply: Our objective to measure the on-off ratio is to quantify the expected idling errors and evaluate it within the context of larger circuits or error correction circuits. Indeed, we are not making any claims that this error is vanishingly small (since, as the reviewer states, cross-talk can be exactly nulled out in some architectures), but rather that it is comparable or lower than our two-qubit gate Pauli errors.

I. SIMULATIONS TO QUANTIFY THE QEC BENEFITS OF SWS GATE NOISE STRUCTURE

In this section, we supplement our referee responses with additional surface code simulations. In particular, we use these simulations to justify our answers to the following questions raised by several referees:

1. Can we delay all erasure checks to the end of a stabilizer while maintaining the QEC benefits of erasures? (Even when leakage can propagate through up to four CZ gates before detection?) See I A and I B

2. What is the benefit of structured errors in the SWS gate in the context of QEC, even if the total gate error is similar to other experimentally demonstrated two-qubit gates? See I C

3. What is the predicted QEC scaling compared to e.g. a transmon-based surface code? (And how is this impacted by imperfect erasure checks?) See I C 4

These simulations use the Stim library along with modifications to accurately model and decode erasures (detected leakage errors) that occur during our gate. We plan for these simulations and new methodology to form the basis of a future, separate publication which focuses solely on the QEC benefits of a dual-rail surface code. We have added two additional appendices (P and Q) to the manuscript which aim to answer questions 1 and 2 as succinctly as possible, since we believe these will be of interest to readers, without delving into the technical details of how these simulations were performed. We reiterate that the main focus of our manuscript is on the hardware details of our two-qubit gate implementation and how we quantify its error rates. In this document, we give a more in-depth overview of our simulation results, as they answer a few key concerns from the referees.

A. Delayed erasure checks preserve threshold and distance scaling

As pointed out by several referees, we assert that erasure checks can be delayed to the end of a stabilizer measurement round in the surface code, without significant loss of threshold or scaling with code distance. Here we provide evidence from simulations that this is true, for the leakage propagation model of our SWS gate (along with insights as to why we expect this result). As a consequence, erasure checks no longer need to be performed immediately after every CZ gate, and we feel their demonstration in this hardware platform is not necessary for the experimental demonstration of a compelling two-qubit gate between dual-rail cavity qubits.

Previous erasure qubit simulations by other authors have not been able to preserve the threshold or scaling with code distance when delaying erasure checks to the end of the stabilizer measurements because they use a leakage model that does not accurately reflect the noise properties of our SWS gate. We have shown with experimental results in our manuscript (See Fig. 4 in our manuscript) that the following properties hold true:

1. If any dual-rail qubit is leaked (to $|00\rangle$ vacuum state) before a CZ gate, the gate is skipped entirely (replaced with identity operations on *both* qubits involved in the gate).

2. If any dual-rail qubit leaks to vacuum during the gate, this can be accurately modeled (see the updated Appendix F) as a 50-50 mixture of performing the complete CZ gate, or performing identity, then followed by leakage jump operators. This means we only need to model leakage happening at discrete locations *in between* CZ gates in our simulations.

We have verified these properties extensively in our manuscript (See Fig. 4) for dual-rail cavity qubits, and they are also expected to hold for other erasure qubits (e.g. atom loss or decay from excited Rydberg states in neutral atom platforms). The requirement for this leakage model to be valid is that we must gradually gain an entangling phase on just one of the basis states to realize a CZ gate, and that leakage mid-way through the gate immediately 'shuts off' the entangling interaction, as is the case for a dispersive entangling interaction.

In order to accurately model these leakage propagation properties, we must remove CZ gates from our stabilizer measurement circuits whenever leakage occurs. With delayed erasure checks, a single leakage event may cause up to four CZ gates to be removed during a stabilizer measurement cycle. This can be thought of as a highly time-correlated Clifford error. However, we can use such a correlated error to our advantage when decoding error syndromes. We have found that if we ignore this property, and instead approximate this noise as a Pauli error channel (such as in [arXiv:2408.00829]), or by using the Pauli twirling approximation, results in a loss of threshold and scaling with code distance.

In our simulations, which are based on Stim, we remove CZ gates from the stabilizer circuits (with different CZ gates removed on each simulation shot, since the 'trajectory' of qubit leakages will differ on each shot). We do not go into the fine details of how these simulations are implemented since this is the subject of a future publication and deserves detailed explanation. We do however briefly summarize the error model used for Section I A in words:

1. Only the CZ gates are error-prone

2. All erasure checks are perfect and are performed on all qubits at the end of each stabilizer round only, which faithfully detect all leakages followed by perfect reset. (Later, we will model imperfect erasure checks.)

3. If a qubit is leaked before a gate, all gates involving this qubit are replaced with identity (until the qubit is reset back into the $|01\rangle$ state following an erasure check).

4. Leakage can only happen before or after a CZ gate.

5. The error model is parametrized by the single error parameter p, whereby:

   - the ancilla qubit leaks with probability $p/4$ before the CZ gate.
   - the ancilla qubit leaks with probability $p/4$ after the CZ gate.
   - the data qubit leaks with probability $p/4$ before the CZ gate.
   - the data qubit leaks with probability $p/4$ after the CZ gate.

To run these simulations, we must also modify the decoding graph depending on the outcomes of our erasure checks on each shot. A clue as to why this error model preserves the distance scaling is apparent in the fact that leakage during the stabilizer cycle does not cause any

'bad hook errors.' For example, when measuring a Z-type stabilizer in the surface code with four CZ gates, leakage mid-way through on the ancilla dual-rail can only propagates errors to the surrounding four data qubits from the set $\{\mathbb{1}, Z_1, Z_1Z_2, Z_1Z_2Z_3, Z_1Z_2Z_3Z_4\}$ which can be written more succinctly as $\{\mathbb{1}, Z_1, Z_1Z_2, Z_4\}$ by multiplying by the stabilizer $Z_1Z_2Z_3Z_4$, which acts as identity on the codespace. The only correlated Pauli error is of the form Z_1Z_2, which is not a hook error since it is always aligned perpendicular to the strings of qubits that define the logical Z operator in the surface code. (For a similar reason, X errors on ancilla qubits halfway through the stabilizer measurement cycle also cause the same benign Z_1Z_2 error, which is why the surface code fairs well with circuit-level depolarizing noise compared to other 2D planar codes). At a physical level, these are correlated dephasing errors affecting our data qubits. However, Hadamard gates on the data qubits will transform these errors such that when measuring an X-type stabilizer, our error set becomes $\{\mathbb{1}, X_1, X_1X_2, X_4\}$ (the error set is unchanged for Z-type stabilizer measurements.)

In fact, when we iterate over all possible locations of leakages and their error syndromes, we find no example of bad hook errors. This holds true for both data and ancilla qubits, and for when all leakage detection is delayed to the end of a stabilizer round. In other words, each detected erasure still only reduces the 'remaining code distance' to Pauli errors by one, and so we expect the erasure scaling with code distance to be preserved i.e. $p_L \propto (p_e)^d$, where p_L is the logical error rate of the stabilizer code, p_e is the per-gate erasure rate and d is the code distance.

This is supported the threshold plot in Fig. 1 , where we show the logical error rate per stabilizer round for a surface code memory (averaged over the X and Z basis), while varying the error parameter p. We plot this for different code distances in order to extract the threshold. More importantly, well-below threshold we can fit our simulated data to the function

$$p_L = B\left(\frac{p}{p^*}\right)^{\alpha d+\gamma} \tag{1}$$

Where B is the (per round) logical error rate at threshold, p^* is the effective threshold for the noise model, α is the scaling with code distance and γ is the offset to this scaling. For Pauli noise, we expect $\alpha = \frac{1}{2}$ and $\gamma = \frac{1}{2}$. A well-known result for circuit-level depolarizing noise is that $p^* \approx 1\%$ and $B \approx 3\%$. For erasure noise, we expect $\alpha = 1$ and $\gamma = 0$ if we are to preserve the distance scaling.

From our simulated data well-below threshold, with the simple erasure noise model described previously, the fitted parameters we extract are $\mathbf{p^* = 5.6\%}$, $\alpha = \mathbf{0.93}$, and $B = 7.1\%$ (where we have fixed $\gamma = 0$). These results are strong evidence that delaying erasure checks to the end of the stabilizer round still results in an impressively high threshold for erasures, and crucially also preserves the distance scaling. For reference, the error model used in Y. Wu et al., Nature communications 13 (1), 4657 which assumes perfect erasure checks after every CZ gate, and that leakage causes a depolarizing error on the other qubit involved in the CZ gate obtained a threshold of $p^* = 5.13\%$, very similar to the value we obtain with our delayed erasure checks (We attribute our slightly higher threshold to be a consequence of the lower noise entropy: leaked qubits can at worse dephase non-leaked qubits rather than cause depolarizing errors, as in their error model.)

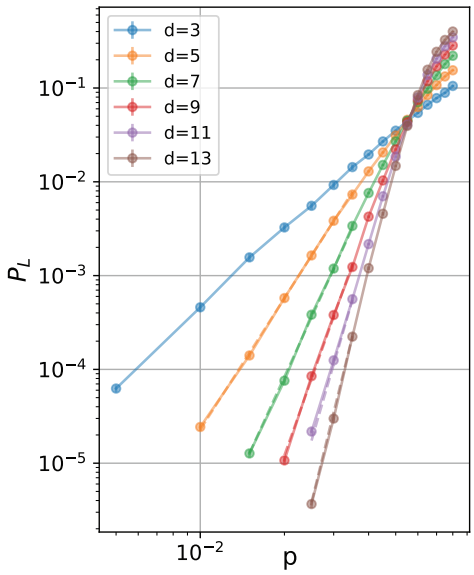

FIG. 1. Simulated surface code memory experiment, averaged over the X and Z basis, under an error model that includes only erasures in CZ gates and perfect erasure checks that are delayed to only occur at the end of each stabilizer round. Probability of CZ erasure per gate is p. Dashed lines are fits to Eq. 1 from which we extract a threshold of $p^* = 5.6\%$ and $\alpha = 0.93$ (from the fitting function $p_L \propto (p/p^*)^{\alpha d}$, showing that delaying erasure checks to the end of each stabilizer round still preserves the favorable error-correction properties of erasures, under our 'leakage skips gates' error model.

B. The effect of imperfect erasure checks

The previous simulation assumes perfect erasure checks on all qubits in the surface code. Realistically, erasure checks are also error-prone. Here we consider the effects of false positives (erasure check declares 'leaked' when qubit is actually in the codespace, resulting in an unnecessary qubit reset), false negatives (erasure check declares 'ok' when the qubit is actually leaked, and hence remains leaked in the next round), and additional erasures induced by the check itself (due to the finite measurement time). We show that these errors can be on the $\sim 1\%$ level, while still maintaining good QEC performance. These are similar to the error rates shown for dual-rail cavity qubit erasure checks demonstrated in other works and hence we do not believe the demonstration of an erasure check is necessary in this work as well.

To quantify the impact of imperfect erasure checks, we observe how p_L reduces with code distance (comparing perfect vs. imperfect erasure checks). The main figure of merit is Λ, the factor by which p_L drops every time the code distance increases by two. We use this definition (the same used in Google's surface code scaling experiments such as Google QAI, Nature volume 638, (2025)) to allow straightforward comparison to other Pauli error models and other experimental platforms, which are typically only investigated at odd code distances. Compared to threshold plots, where the entire error model must be parametrized by a single error rate, the Λ metric allows us to compare error models with several independent error channels. All Λ plots referenced here are shown in Fig. 5 of Section I C 4.

For a 'Λ-plot', we must fix a set of error parameters and vary only the code distance. The noise model we use for these simulations is described below:

1. Only the CZ gates and erasure checks are error-prone
2. All erasure checks are performed on all qubits at the end of each stabilizer round only, if leakage is declared the qubit is perfectly reset.
3. If a qubit is leaked, all gates involving this qubit are replaced with identity.
4. Leakage can only happen before or after a CZ gate, or during an erasure check.
5. The error model is parametrized by the error set $\{p_e^a, p_e^d, p_{\text{fp}}, p_{\text{fn}}, p_e^{\text{EC}}\}$, where p_e^a is the probability the ancilla (control) qubit leaks during the CZ gate, p_e^d is the probability the data (target) qubit leaks during the CZ gate, p_{fp} is the probability of a false positive error in the erasure check, p_{fn} is the probability of a false negative error in the erasure check, and p_e^{EC} is the probability of leakage occuring during the erasure check. The error model can then be described as
   - the ancilla qubit leaks with probability $p_e^a/2$ before the CZ gate.
   - the ancilla qubit leaks with probability $p_e^a/2$ after the CZ gate.
   - the data qubit leaks with probability $p_e^d/2$ before the CZ gate.
   - the data qubit leaks with probability $p_e^d/2$ after the CZ gate.
   - the qubit leaks with probability $p_e^{\text{EC}}/2$ before the erasure check
   - the qubit leaks with probability $p_e^{\text{EC}}/2$ after the erasure check
   - if a qubit is not leaked, the erasure check declares 'leaked' with probability p_{fp}
   - if a qubit is leaked, the erasure check declares 'ok' with probability p_{fn}

The error rates chosen for the CZ gate in this simulation reflect the values measured for erasure rates in the SWS gate from our repeated quantum state tomography experiments (Fig. 3a), as shown in the table below:

Error Parameter	Perfect erasure checks	Imperfect erasure checks
p_e^a	0.4%	0.4%
p_e^d	0.1%	0.1%
p_e^{EC}	0	1%
p_e^{fp}	0	1%
p_e^{fn}	0	1%
Λ	284	22

TABLE I. Error rates used for the Λ plots for Section I B. In the bottom row we show the value of Λ extracted from our fits and plots shown in Fig. 5

The values for Λ are extracted from the simulated data below. Although imperfect erasure checks do significantly degrade our error correction scaling performance, the value of Λ remains impressively high (Another way to interpret Λ is that $\Lambda = 20$ means our surface code performs equivalently to a Pauli-error-only model that is ×20 below its threshold)

C. Comparing structured CZ gate Pauli noise to other noise models

We assert in the manuscript that having structured errors in the CZ gate $p_e > p_Z > p_X$ (and additional asymmetry between the control and target qubits) can greatly benefit quantum error correction. Here, we present simulations that hold the total error rate per gate constant, but show how noise structure can give higher thresholds and better QEC scaling performance. In these simulations we assume the CZ gate is the only source of error, so we can isolate the benefits of different parts of our noise structure. For these comparisons, we employ threshold plots since there is a single error rate that parametrizes each error model.

1. *Circuit-level depolarizing noise*

Our reference noise model will be the standard circuit-level two-qubit depolarizing noise applied after every CZ gate with strength p (we apply the channel $\varepsilon(\rho) = (1-p)\mathbb{1}\rho\mathbb{1} + \frac{p}{15}\sum_{i,j}\sigma_i\sigma_j\rho\sigma_i\sigma_j$, where $\sigma_i = \{\mathbb{1}, X, Y, Z\}$). From this we obtain a threshold of **1.1%**, using the default Minimum Weight Perfect Matching algorithm in the PyMatching library. The noise model is used very commonly to model 'unstructured' noise in two-qubit gates for many qubit platforms such as transmons. The threshold plot is shown below.

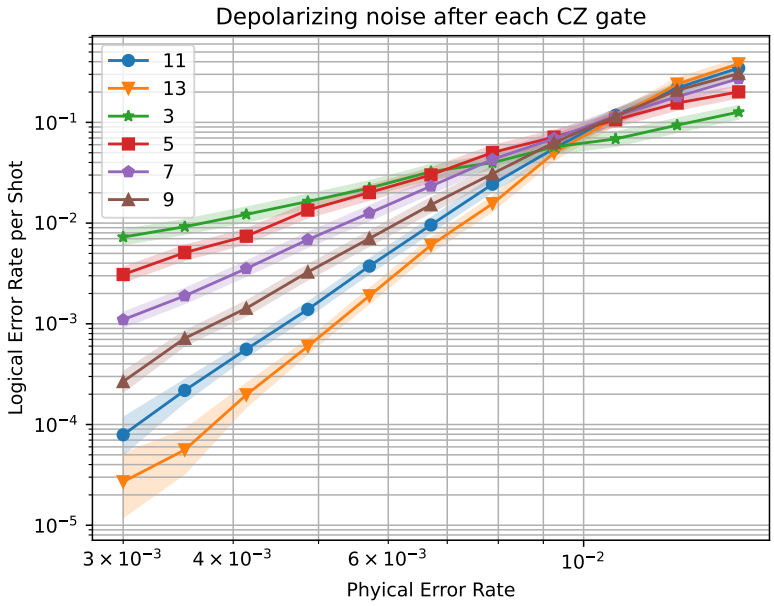

FIG. 2. Threshold plot for an error model that only includes two-qubit depolarizing noise after every CZ gate. As expected, we obtain a threshold close 1%, which is a well-known result for completely 'unstructured' Pauli errors in a two-qubit gate.

2. *Dephasing only (balanced)*

The next noise model we consider is one where only dephasing errors affect our CZ gate, for which we apply the error channel

$$\varepsilon_Z(\rho) = (1 - p_Z^c - p_Z^t - p_{ZZ})\mathbb{1}\rho\mathbb{1} + p_Z^c Z_c\rho Z_c + p_Z^t Z_t\rho Z_t + p_{ZZ} Z_c Z_t\rho Z_c Z_t \tag{2}$$

For the 'balanced' dephasing simulations, we set $p_Z^c = p_Z^t = p/2$ and $p_{ZZ} = p^2/4$. We do not include any other errors, since we want to highlight the error-correction benefits that arise from the structure of our Pauli noise in CZ gates. From our threshold plots, we obtain a threshold of **2.2%**, which is significantly higher than the two-qubit depolarizing noise model, illustrating the importance of structured Pauli noise on the threshold, even when the single qubit gates do not preserve the Pauli noise bias.

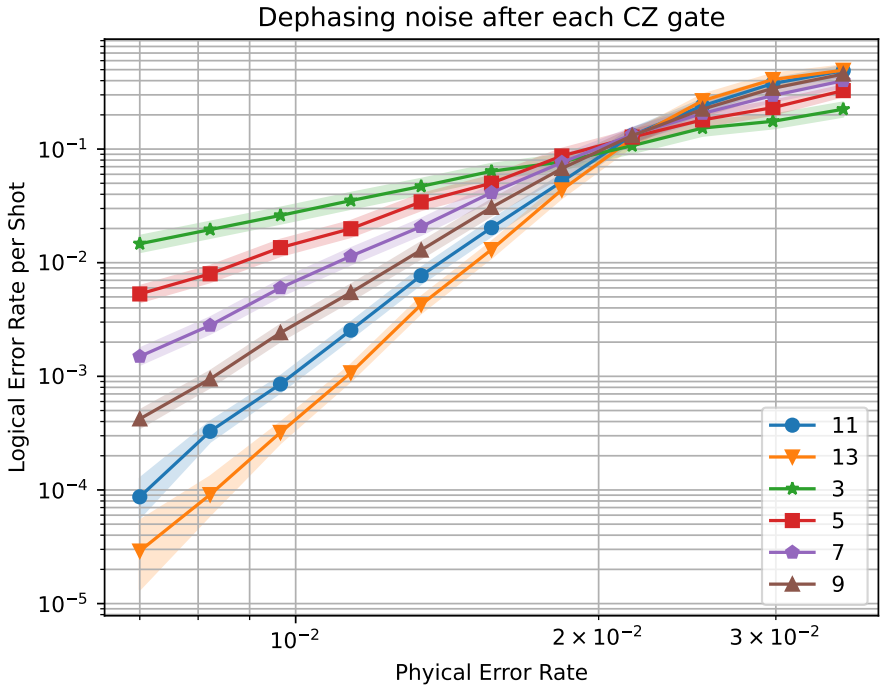

FIG. 3. Threshold plot for an error model that only includes uncorrelated dephasing noise after every CZ gate. The ancilla and data qubit are equally likely to suffer from dephasing under this error model. The increase in threshold to 2.2%, is about double the value obtained from a depolarizing noise model, and indicates how structured Pauli noise can improve error correction performance (even when the single qubit gates do not preserve this bias.)

3. *Dephasing only (asymmetric)*

This noise model is the same as above, but in order to highlight the benefits of asymmetric noise wherein the ancilla qubit picks up the majority of the dephasing, we set $p_Z^c = 4p/5$, $p_Z^t = p/5$ and $p_{ZZ} = p_Z^c p_Z^t$ from which we obtain an even higher threshold of **3.2%**, as shown in the threshold plot below. The reason for this higher threshold is thought to be related to the fact that dephasing on the ancilla qubit during a CZ gate will propagate through as a measurement error in the stabilizer measurement result, and will not propagate onto any data qubits. A surface code memory is good at tolerating incorrect stabilizer measurements whereas Pauli noise on the data qubits is more damaging since it must be correctly identified by subsequent stabilizer measurements. Although the error rate is still p per CZ gate, we find that the structure of the Pauli noise has a significant impact on the error threshold, which is why we say our gate 'does better' than other gates even if the total Pauli noise is still $\sim 0.1\%$

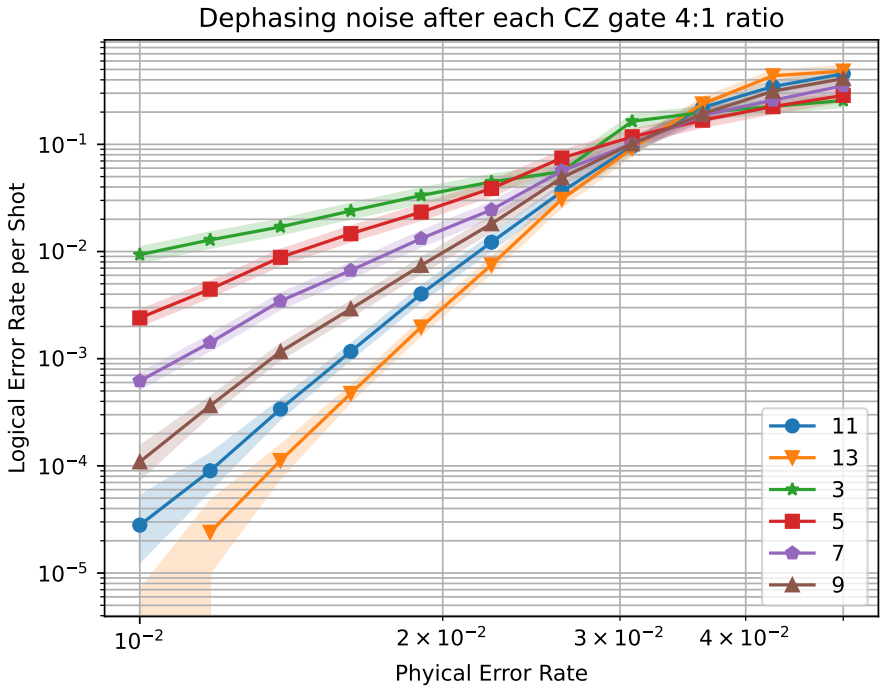

FIG. 4. Threshold plot for an error model that only includes uncorrelated dephasing noise after every CZ gate, but when the ancilla qubit suffers from more dephasing errors than the data qubit in a 4:1 ratio (same total dephasing rate per gate as Fig. 3). The threshold further increases to 3.2%, indicating how structured *dephasing* noise where the ancilla qubit suffers from more errors than the data qubit can be leveraged for a substantial improvement in error correction performance.

4. *Error correction gain plots (Λ-plots)*

Whilst threshold plots are useful for showing the benefits of Pauli noise structure, we use "Λ-plots" to quantify error correction performance when both Pauli and erasure errors are present, and to investigate behavior well below threshold. Here we fix the error rates for our error model and instead vary the code distance to extract Λ. For our SWS gate simulation we use the parameters shown in the table, and also presume imperfect erasure checks. Now our error model has both Pauli and erasure-type errors. In our simulations, the Pauli error channels only act on qubits that are still in the qubit, after each CZ gate operation. We

Error Parameter	Structured CZ noise
p_Z^c	0.04%
p_Z^t	0.01%
p_{ZZ}	0.01%
p_X^c	0.0003%
p_X^t	0.00005%
p_e^a	0.4%
p_e^d	0.1%
p_e^{EC}	1%
p_e^{fp}	1%
p_e^{fn}	1%

TABLE II. Error rates used for the structured noise model in the red plot in Fig. 5 . Error rates for the CZ gate are the same as those measured in the manuscript. The false positive and false negative, and induced erasure rates due to imperfect erasure checks are all set to 1%, a rather conservative value.

show Λ plots where we compare this noise model vs depolarizing noise models and the SI1000 noise model, both with $p = 0.1\%$. The point of these plots is to show that even though our total CZ gate error rate is around 0.5% (and a total Pauli error rate of 0.06%), we obtain a more impressive value of Λ compared to the unstructured noise models.

These simulations are by no means complete, since we have not yet included the effects of reset errors, single qubit gate errors, or idling errors, which is the subject of future work. However, these simulations aim to compare Pauli noise and our structured noise on roughly even footing in order to illustrate the benefits of the structured noise in our SWS gate vs. unstructured noise.

Finally we compare noise models against each other from the data presented in Fig. I C 4, where each comparison helps inform us about a different aspect of SWS error structure in a surface code context.

- (green) vs. (purple). If we assume errors only come from the CZ gate, then a surface code based on our SWS gate with the noise values we have presented in the manuscript would have a Λ that is roughly double that of a standard CZ gate with depolarizing noise of 0.1%. Using the Λ metric was the most direct comparison we could think of for comparing the SWS gate to more standard two-qubit gate error models, as per comment A.6.

- (blue) vs. (orange) vs (red). In orange, we include the effect of imperfect erasure checks, with 1% error rates for the false positives, false negatives, induced erasures. This tells us two things. First, in a QEC context, the effect of imperfect erasure checks cannot be ignored and ends up being a significant source of logical errors related to erasure-like error channels. However, Λ still remains rather impressive, even with imperfect erasure checks. The inclusion of additional CZ dephasing errors (green) roughly halves Λ, suggesting that Pauli errors in the gate and erasure-like errors contribute similarly to the overall logical error rate.

- (purple) vs. (brown). Two qubit gate error is often the single most dominant error channel for QEC. However, an accurate prediction of Λ requires modelling many other independent error channels, such as measurement error, single-qubit gate error and idling errors. Such errors are included in the 'SI1000' noise model, which is commonly used by Google Quantum AI and others working with transmon-based QEC approaches to model their qubits. For the surface code, it has a threshold of 0.3-0.5% (See Gidney et al., Quantum 5, 605 (2021)), much lower than the 1% threshold for standard depolarizing error, which suggests that an accurate prediction of Λ cannot come from two-qubit gate performance numbers alone. Although the SWS gate looks promising as an source of low two-qubit gate errors, full accounting of multiple error channels is needed to predict an experimentally realistic value of Λ, which is the main focus of our upcoming publication.

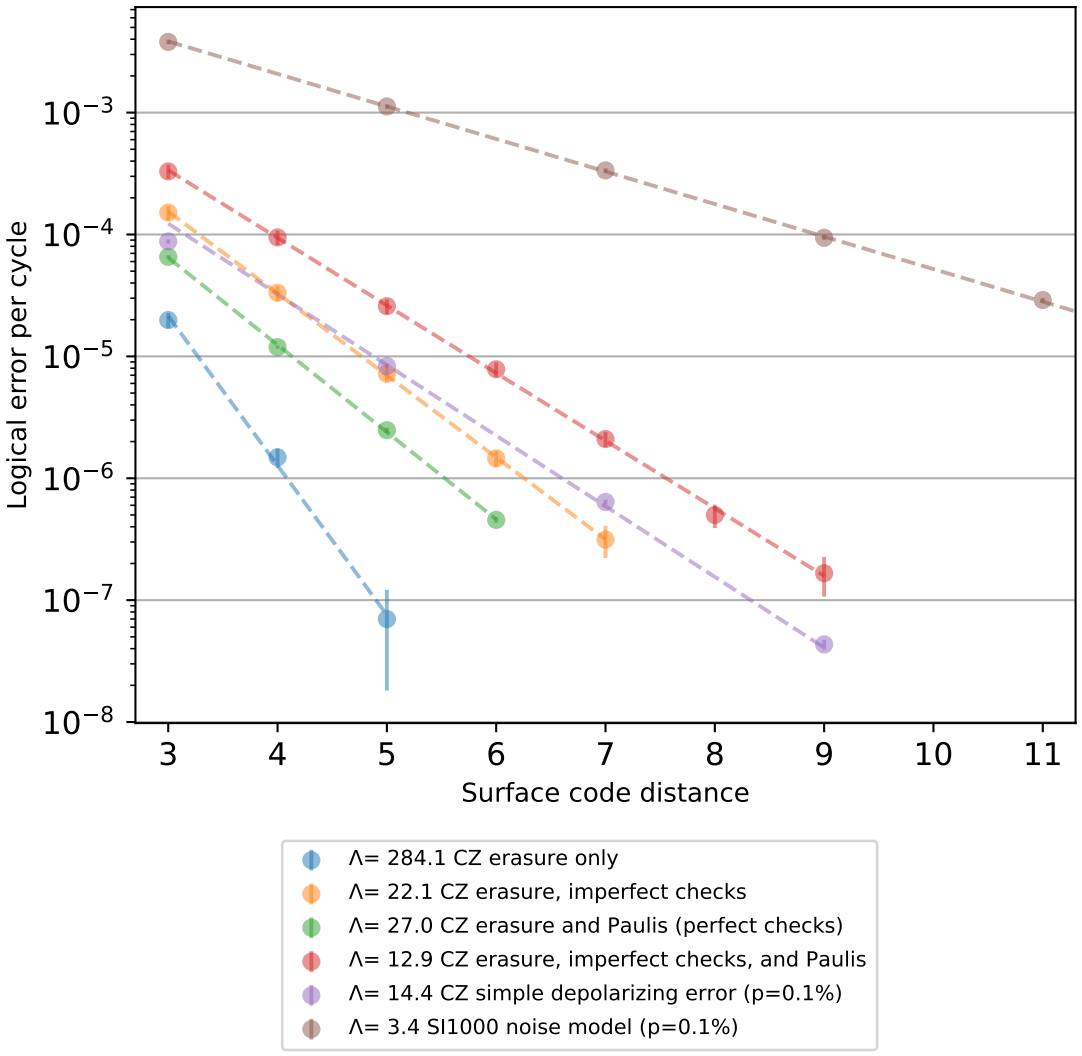

FIG. 5. Various "Λ-plots" under different noise models. The simualted logical error rate per round of a surface code memory, averaged over the X and Z logical bases, is shown as a function of inreasing surface code distance. Λ is defined to be the suppression in logical error rate (per round) when the code distance is increased by two. Different insights can be gained by comparing pairs of error models against each other. For instance, simple depolarizing noise (purple) vs SI1000 (brown) indicates very different values of Lambda even though the 'strength' of the noise is the same. In actual quantum hardware, where error rates can approach 0.1%, but many operations (such as measurement, idling etc.) are also error-prone, we obtain a Λ that is about 4× worse than the simple depolarizing error model, and shows how daunting achieving a $\Lambda \sim 10$ can be in practice for the surface code.

Referee responses and summary of revisions, Round 2
Manuscript number: 2025-05-11044

(Dated: June 5, 2026)

OVERVIEW OF RESPONSES

REVIEWER 1

Reviewer Comment A.1 —

In the previous round of review, we mainly expressed our concern on the missing of mid-circuit erasure check, the high SPAM error rate and the related scalability issues, and the QEC performance comparison between the structured noise (erasure dominant + Pauli Z bias) and a high-fidelity generic noise model. We appreciate the authors' enormous efforts on the clarification of the major focus of this work (engineering and benchmarking this novel 2-qubit gate with erasure-dominant error structure), the explanation for the origins of the SPAM error, as well as a detailed simulation on the QEC performance under the structured noise model.

We think the results of this work is fairly strengthened after this revision and we are convinced that the novel error structure engineered from the dual-rail CZ entangling gate could be beneficial for the QEC performance compared with the general depolarized noise model, provided that the SPAM and mid-circuit measurement fidelities are further improved.

Reply: We thank the reviewer for the positive outlook on the new version of our work. We appreciate the comments from all reviewers and their input have greatly improved the manuscript. We also agree that the dual-rail error structure can be highly beneficial for QEC.

Reviewer Comment A.2 — On the other hand, despite that the authors argue that there is no need to perform mid-circuit erasure check after every CZ gate in the QEC setup, it is still a necessary gadget after every stabilizer measurement round, and therefore the characterization of its performance is also important in order to make QEC advantage claim for the dual-rail framework. We checked one related paper mentioned by the authors [S.de Graaf, S. Xue et al, npj QI 11(1) (2025)] and found out that the current infidelities for the erasure check (3-4% erasure or false-negative) is still a little higher than what is used in the QEC simulation. The 0.3% Pauli error from the check is also not included in the simulation. Therefore, we think there is still sufficient evidence to believe that, at the current stage of the hardware performance, it is hard to make any claim that the dual-rail framework could provide QEC advantages compared with other architectures based on superconducting qubits.

Reply: We thank the referee for this comment, which concerns the need to characterize the impact of faulty erasure checks at the end of each stabilizer round on QEC performance. We study this effect by adding depolarizing noise to the mid-circuit erasure checks used in the previous simulation (red plot in Fig. 5 in the first referee response). The results of which are shown in Fig. 1, and show that we want to target a depolarizing error per check $\lesssim 0.1\%$ in order to maintain a Lambda scaling close to 10.

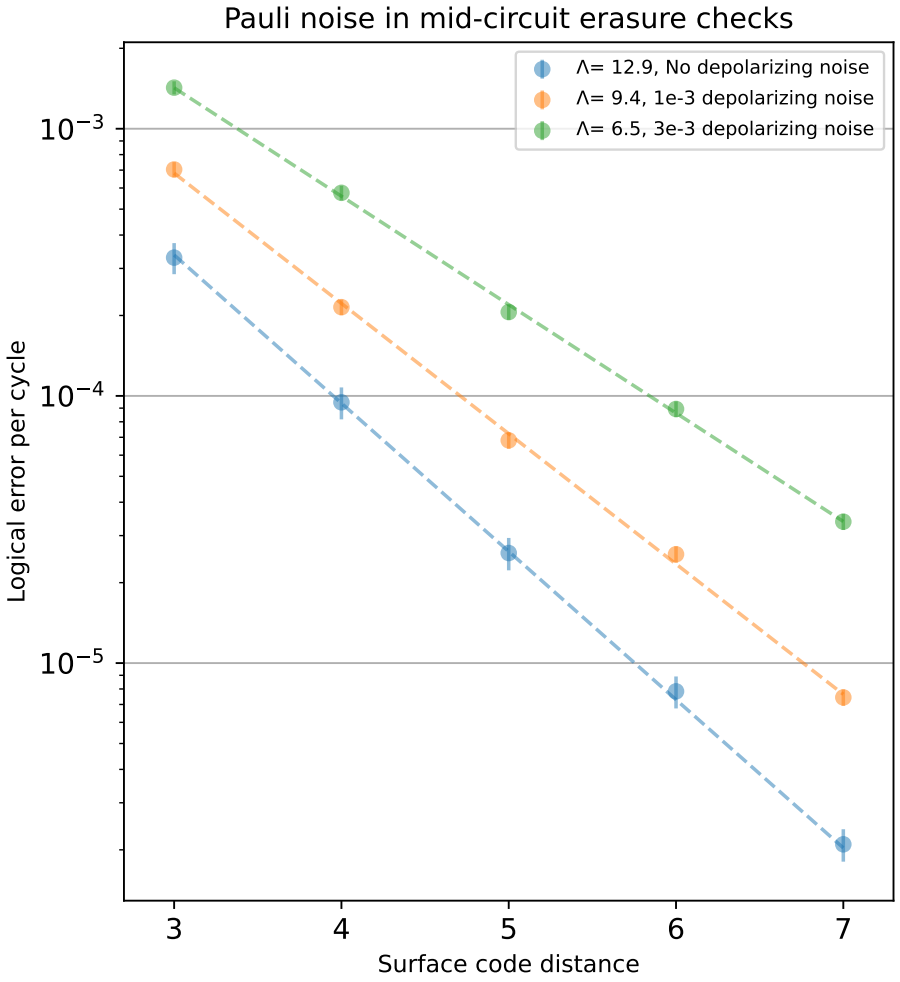

FIG. 1. The effect of Pauli noise in erasure checks on surface code performance. The noise model is the same used for the red plot in Fig. 5 of the first referee response (imperfect erasure checks, erasures in CZ gates, Pauli errors after 1Q and CZ gates), with the addition of a single qubit depolarizing error channel applied after each erasure check on the data qubits.

Whilst an erasure check for dual-rail cavity qubits with a Pauli backaction $< 0.1\%$ has not yet been published in the literature, we are confident there are ways to achieve $< 0.1\%$ Pauli backaction with the scheme used in [S.de Graaf, S. Xue et al, npj QI 11(1) (2025)], since this work was largely coherence limited by lower-than-typical cavity coherence times.

Reviewer Comment A.3 — In the meantime, while we agree that "for the purpose of this work" (benchmarking CZ gate) the high SPAM error is acceptable, we still keep our concern that it will become a problem when people want to scale up the system in the future. It will be good to see if this can be reduced to 1%-level while maintaining the gate performance for future hardware.

Reply: We concur with the reviewer on this matter and have added additional details in SI section I subsection B on mechanisms currently limiting the erasure assignment during EOL measurements

and the path to achieve $\sim 1\%$ false positive erasure rate in our EOL measurements.

REVIEWER 4

Report on Revised 2025-05-11044 Summary

The authors have clearly addressed most of the questions raised by us and the other reviewers. In addition, the revised manuscript includes two significant claims.

- Mid-circuit erasure checks are not necessarily vital during the error correction cycle in the proposed dual-rail architecture.
- The dual-rail architecture and the CZ gate presented in the paper can, in principle, improve distance scaling compared to a standard two-qubit gate of similar fidelity.

These claims were substantiated through numerical simulations in the manuscript, and we also found a corroborating paper [PRX Quantum 6, 040354 (2025)] that is relevant to the first point. The importance of these findings extends beyond the creation of novel, high-fidelity two-qubit gates and deserves attention from the broader community. Whether it belongs in Nature or a more specialized, but still very high-impact, journal like Nature Physics is an editorial decision. Minor technical points The manuscript needs more explanation regarding the large false-positive error rate (6% according to the Appendix). False-positive errors have a nonnegligible effect on the error threshold under erasure checks, and 6% may not be tolerable for error correction, as relevant experiments [PRX 14, 011051 (2024)] or numerical simulation [PRX Quantum 6, 040354 (2025)] target an error of 1% or less, including the current manuscript. From the response, the authors seem to be aware of this issue, but we believe paper should clarify the conditions that need to be satisfied to achieve the targeted 1% false-positive error rate.

Reply: We have added additional details in SI section I subsection B about mechanisms currently limiting the false positive assignment to easure during our EOL measurements, additionally we have also outlined a path to get this error down to $\sim 1\%$ level.

Reviewer Comment B.1 — In Figure 2c, which illustrates Bell-state tomography, the authors speculate that the quadratic decrease in fidelity and purity at larger gate depth may result from drifts in calibration parameters or fluctuation in frequency.

However, it is not clear that parameter drift would be sufficiently pronounced over the timescale of these specific measurements to cause visible effects. Could there be other plausible explanations such as, for instance, the accumulation of erasure errors during the CZ gate that results in dephasing or an increase in undetected erasure error events?

Reply:

We have considered the possibility of accumulation of erasure and undetected leakage by our End-of-line (EOL) measurements as sources of the non-linear fidelity with gate number but have determined the effect is too small to be the main source of the quadratic decrease, given our observed error rates. We arrive at this conclusion for several reasons. Firstly, the "leakage-skips subsequent gates" property applies to both our single and two qubit gates. If a dual-rail qubit undergoes photon loss to the vacuum state at any point in the circuit, it will remain in the

vacuum state until it is measured at the end. We have not observed any so-called "seepage errors" where leaked dual-rail qubits re-enter the codespace and potentially present themselves as Pauli errors. It is expected that seepage can occur when a dual-rail photon loss is followed by one of the cavities gaining a thermal excitation. However, given the T_1 and low thermal population of our cavities, we expect this to be exceedingly rare, similarly rare to our observed bitflip rates. A related mechanism, which we term "apparent seepage", occurs when we mis-assign a leaked qubit to be in the dual-rail codespace in our EOL measurement. However, the probability of this misassignment error is small ($< 1\%$), and is not expected to increase with gate depth, so this is not expected to produce a quadratic decrease.

One mechanism that could explain this quadratic decrease is parameter drift on the time-scale of < 1 ms, in particular fluctuations in the power of our parametric drive that enacts the cavity-coupler beamsplitter. This operation is very sensitive to the drive power of the pulse since we have a large ratio ($\sim$10) between the Stark shift on the coupler and the beamsplitter rate of the operation. This Stark shift results in a Z phase update on the control qubit. Loosing track of this phase results in dephasing that compounds with gate number, and can be thought of as a source of low frequency dephasing noise.

Reviewer Comment B.2 — There is a small typo in Appendix N: $|0\rangle\,|1\rangle$ should be "$|01\rangle$ state" for consistency. Also, in the below sentence, "In contrast, the evolution of the $|01\rangle$ dual rail" should be $|00\rangle$ instead of $|01\rangle$. On page 5, "In Appendix K we performed repeated IRB" should be "In Appendix L."

Reply: Thank you for the detailed review and we have addressed this typo.