nature portfolio

Peer Review File

Hardware-efficient quantum error correction using concatenated bosonic qubits

Corresponding Author: Mr Harald Putterman

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 authors demonstrate quantum error correction (QEC) on a chain of 5 cat-qubits. Bit-flips are corrected by two-photon dissipation at the level of each cat-qubit. Phase-flips are corrected by an outer repetition code that measures the stabilizers XX for neighboring qubits. The entire QEC cycle is fast enough and introduces sufficiently low errors, that the logical phase-flip errors diminish when scaling the code from distance 3 to distance 5.

In my opinion this experiment is a landmark in the field of QEC.

Errors are the main roadblock towards the emergence of useful quantum machines. QEC provides a solution, in principle. However, currently investigated QEC architectures require daunting hardware overheads (e.g. surface code of transmons). There is a pressing need for new ideas that reduce this hardware overhead. Cat-qubits are one path forward (amongst many other promising ideas). The outstanding challenge is that although the entire QEC architecture is expected to be composed of a smaller number of qubits, and therefore easier to operate, each qubit is much more complicated than a transmon. There is therefore a compromise to find between (i) the complexity of the building block (transmon vs. cat-qubit ?) and the (ii) the complexity of the QEC architecture (surface code vs. repetition code ?). The experiment by Putterman et. al. shows that they are capable of mastering the fabrication and control of arrays of cat-qubits to a level where QEC actually works. Although these exotic dynamical qubits are still in their infancy, this rapid pace of progress is extremely encouraging to overcome the huge challenges lying ahead of universal fault-tolerant quantum computation.

The authors have made several bold and risky technological advances that have paid off.
(i) The matched chi ge - ef transmons, for fixed frequency transmons.
(ii) The low Kerr buffer modes by compensating the Kerr due to the cos(2phi) with the Kerr due to the junction array.
(iii) The flip-chip architecture with tantalum cat-qubit resonators
(iv) The tunable couplers between the transmon and cat-qubit modes.
Getting all these ingredients to work together on a chip of 5 cat-qubits, their 5 buffer modes and 4 ancillary transmons, is a true technological "tour de force".

I particularly appreciate that the authors are explicit on technical details like fabrication, fridge wiring and crosstalk cancellation. The peer-reviewing process in scientific journals is a gem to be protected. By supplying detailed information, referees (as well as the entire scientific community) may check the consistency of the results, challenge the conclusions and go as far as reproducing the experiment when necessary. In my opinion, it is important that all contributors to scientific journals (including companies) abide by editorial rules, and do not compromise the rigor of the reviewing process.

Before delving into more detailed and technical questions and comments, let me say upfront that I recommend this paper for publication in Nature. It is well written and the figures are of highly quality.

Prof. Zaki Leghtas.

****

Detailed comments and questions:

1. In the abstract this sentence is misleading "We realize a noise-biased CX gate which ensures bit-flip error suppression is maintained during error correction". This may come across like a CX between two cat-qubits. Instead, what is performed is a CX between a cat-qubit and a transmon.

2. The authors mention multiple times the "threshold" of their code. They are careful in saying it is the "phase-flip correcting repetition code". For completeness they should however add a sentence in the text about the fact that overall (including bit-flips and phase-flips) this cat-qubit architecture does not possess a threshold. This is discussed in [Phys. Rev. A 103, 042413]. Although this may not be of practical significance, it is a conceptual fact that is worth mentioning.

3. Minimizing the Kerr of the storage mode is crucial since the two-photon dissipation is pulsed (it is in fact off for a large fraction of the QEC cycle). This Kerr is partly induced by the Kerr of the buffer mode, which therefore needs to be minimized. The authors describe in a separate paper [ref 25] that they minimize the buffer Kerr by balancing the Kerr induced by the buffer junction array (negative) and the "cos(2phi)" contribution from the stray inductances in series with the small junctions of the ATS (positive). Could they explain how they fight against the Kerr induced by junction asymmetry? This is random, and difficult to measure at room temperature.

4. In Fig 1b some transmission lines cross. How was this achieved? Are there suspended bridges on the chip?

5. In Fig 1f, I would suggest plotting 1/Tphaseflip (instead of Tphaseflip) in order to visualize the linear scaling as a function of alpha^2. The y-ticks could be in "us^(-1)" in order to immediately read-off Tphaseflip.

6. In Fig 1e, there are missing ticks on the Wigners, and a missing colorbar.

7. Combining the chi matching approach with two-photon dissipation where the latter corrects mis-rotations is a brilliant idea. This leverages the power of two-photon dissipation: it "cools" the system down to the cat-qubit quantum manifold. This solves the problem of leakage and error accumulation.

8. Fig 2b: Is the Wigner plot post-selected for instances where the transmon was measured in g 10 times? i.e. is the smearing of the coherent state only due to Kerr or also to chi rotations due to the transmon jumping to e?

9. Fig 2c. inset. Again, I would prefer the inverse of phase-flip time vs alpha^2 to verify the linear scaling. What is the solid line?

10. Fig 2c. orange curve. Have you tried to post-select on having measured the transmon in g all the time? I would expect that the orange curve should then climb up to meet the black curve. The number of instances where you measure g at every shot over 10s of ms might be very low, in which case you could for e.g. progressively remove instances where you measured the transmon in e (or f) and see that depending on how many trajectories you throw away the orange meets the black.

11. Fig 2a,b. The colorbars are missing.

12. Is CX^2 simply waiting twice as long as CX during the dispersive interaction of the transmon and storage?

13. Fig 3b: why are the first and last points above the average?

14. Fig 3d: how do you measure pmeas?

15. Fig 3f: the meaning of the "faded points" is unclear to me. Also, they crowd the plot.

16. You mention "we carefully tune the syndrome measurements to avoid parasitic effects that can induce cat-qubit bit-flip errors, including [...] spurious two-level systems (TLSs) in the ancillas". How do you avoid TLS? Is just by thermal cycling until you are lucky enough that no TLS couples to any transmon?

17. About Sec V: since you have access to single shot Z measurements, and that in principle this measurement is QND, have you measured single trajectories for Z? If so, have you looked for correlations between Z trajectories of different cats?

18. Section VI: you write that for the distance 5, the minimal measured eL = 1.65% ± 0.03%. For distance-3 sections you find eL = 1.83% ± 0.03% and eL = 1.67% ± 0.04%. I am assuming the eL = 1.75% quoted in the abstract is the average the two latter numbers. This is a bit misleading, and contrasts with the otherwise very honest and clear claims of the paper. The results of this paper are very impressive, so there is no need to make a controversial claim in the abstract (comparing d=5 to the average of the two d=3 codes, knowing that one of the d=3 codes had the same eL than the d=5). For the abstract, I encourage the authors to highlight the results of Fig3f, where we see that the d=5 outperforms both d=3 codes.

19. The last paragraphs in the conclusion are very instructive. The authors talk about the "performance floor" due to the transmon. Could they be more explicit on the numbers? What if you could make a 1 ms transmon (or Fluxonium), what would that noise floor be? Since you are only sensitive to the double decays during the CX, the numbers might be quite good. It is worth mentioning them, because this strategy could go a long way.

20. There are multiple loops on this chip, could the authors say something about flux drifts: how they minimize them

(shielding, filtering etc) and how they actively compensate them (recalibrations?). How large are these drifts (in units such as mPhi0 per hour)?

Referee #2

(Remarks to the Author)
The authors report on a quantum logical memory experiment, realizing a distance-5 repetition code of bosonic cat qubits, on a superconducting circuit architecture. The cat qubits utilize higher levels of harmonic oscillators to encode the quantum information in (a superposition of) coherent states stabilized by two-photon dissipation, with a large noise bias (bit-flip time much larger than phase-flip time). Capitalizing on this noise bias, the authors can afford to correct only one type of error, therefore requiring only a one-dimensional repetition code, rather than a two-dimensional quantum error correction (QEC) code, therefore justifying the "hardware-efficient" qualifier.
This is a novel result: cat qubits were demonstrated on their own, and planar QEC codes such as the surface code as well, but the repetition code of noise-biased bosonic qubits is a first (to my knowledge). It is also not a trivial result, because of the requirement to preserve the noise bias in the syndrome measurements. The authors cleverly address this using a g-f encoding of the ancilla qubits, matching the dispersive shifts, and detecting erasures when the f-state decays to e.
I also find the experiment to be timely, with many experimental quantum logical memory results appearing in the past few years (Ref 28-35), at or below threshold.

Generally, the data is of high quality and supports most of the claims in the manuscript (besides one, see below). The figures are well-presented, with error bars, and reliable trends.

My two main suggestions for improvements are the following:

1. The claim to be below threshold is a bit tricky, as the repetition code distance indeed improves the phase-flip time (Fig 3f), but decreases the bit-flip time (Fig 4b). What ultimately matters is the logical memory error rate. There the description is suddenly vague, such as when the abstract reads "demonstrating the effectiveness of bit-flip error suppression throughout the error correction cycle." (I am honestly not sure I parse this sentence as the authors intended.)

I believe it would make sense to bring some clarity and be transparent about:
a) Why is the minimal value of |alpha|^2 (=1) also the one that represents the best results in the distance-3 codes? Are the authors sure that going below doesn't improve the performance even further? It seems odd in Fig 5a to have the optimum as the extremum of the explored range.
b) What is the logical error suppression factor when increasing the hardware code size (overall seems to be possibly 1.75/1.65=1.06 with a large error bar)?
c) is the concatenated-cat code beyond any break-even (if the information was instead encoded in a Fock state of one of the harmonic storage mode)? Seems to be just at or around break-even.

Note: I don't believe it would reduce the impact of the paper, rather provide a clean and easily understandable status.

2. The text clarity could be improved at various places (often refers to statements without enough details to follow, yet at times wondering if really required, or could be made much shorter). It gives an overall feeling of being rushed during writing. Some examples:
a) The conclusion has 4 paragraphs, which read:
(i) summary of the work (ok),
(ii) what is limiting and could be done better, *as demonstrated in Manuscript in preparation* (!! is it https://arxiv.org/abs/2409.17556?),
(iii) one could then tile a 2D code anyways instead of 1D, see analysis in *Manuscript in preparation* (!!),
(iv) what could happen hypothetically (in too much details) if the problem of implementing bias-preserving gates between cats is solved.
Paragraphs (ii)-(iv) could be combined in a single much shorted one, that provides the same ideas with more punch and less reliance on (uncertain) future works.

b) "Moreover, the buffer-mode parameters are carefully chosen to minimize other parasitic buffer-induced nonlinearities on the storage mode."
What parameters? What other parasitic nonlinearities? This is a filler sentence that either needs expansion, or be removed completely and just refer to the suppl. mat.

c) "Only higher-order or suppressed ancilla error mechanisms, such as two sequential decay events or heating will cause bit-flip errors on the data cat qubit."
Which mechanisms are considered higher-order or a suppressed? What is the difference? Aren't they suppressed because they are higher-order?

d) "The small degradation relative to the reference performance at |alpha|2 >~ 3 is due to heating events in the ancilla or coupler during the CX^2 gate."
How high is heating? Is a factor ~2 difference at |alpha|^2=4 really a "small difference"?

e) "we show a control sequence involving a CX gate similar to the one used for an error correction syndrome measurement."
Is it not equal? What makes it a "similar" gate rather than the same?

f) "Notably, all these imperfections can be corrected with high probability when the two-photon dissipation is turned back on after the CX gate"
How high are we talking?

g) "Each syndrome measurement cycle has a conservatively chosen duration"
Can you be quantitative on "conservatively"? How much buffer time is planned?

h) "we carefully tune the syndrome measurements to avoid parasitic effects that can induce cat-qubit bit-flip errors, including [...] spurious two-level systems (TLSs) in the ancillas"
I don't see any tuning of syndromes for this. The suppl. mat. only refers to "temperature cycle the dilution fridge".

i) Fig1a and b would benefit from labeling Si, Ai, ...

Referee #3

(Remarks to the Author)
The submission “Hardware-efficient quantum error correction using concatenated bosonic qubits” demonstrates experimentally the use of biased-noise qubits in a quantum error-correcting memory. The physical qubits are bosonic cat qubits, which are notable for their extreme noise bias. It is natural to examine the use of a biased cat qubit in a quantum repetition code, where phase-flips are corrected at the logical level, while bit-flips are suppressed at the physical level.

This submission presents an impressive and timely research result in quantum error correction. Its impact and level of scientific interest will be high, given recent results in surface code quantum error-correcting memories, with which these present results are very complementary. The key technical innovation in this paper is the demonstration of a noise-biased CNOT gate, which underpins the scheme, and this is combined with an impressive integration of a number of previously-demonstrated components and techniques into a system that performs in a regime relevant to fault-tolerant quantum computing. We have not identified any fundamental flaws that would prevent publication. For these reasons, we believe the paper is clearly of the standard that warrants publication in Nature.

The paper is well-written, and the flow of results from the level of physical qubits to the level of logical memory makes the paper easy to read for a broad audience. This comes at the cost of specificity, but this is unavoidable given the length constraints.

We have provided a number of comments and questions for the author’s consideration.

Comments:

In the introduction, “A complementary QEC paradigm is to use a layered approach to noise protection, in which one starts from a qubit encoding that natively protects against different noise channels and suppresses errors.” While recognising that this statement is somewhat ‘setting the scene’, it is not entirely accurate that an approach of concatenating a bosonic code with a repetition code sits in a different paradigm than surface-code-based QEC. We’d encourage the authors to state more directly the differences between this approach and others.

“One example is bosonic qubits, where qubit states are encoded in the infinite-dimensional Hilbert space of a bosonic mode (a quantum harmonic oscillator), and the extra dimensionality of the mode provides redundancies that can be utilized to perform bosonic QEC.” We appreciate the need for brevity, could the authors elaborate on what exactly the redundancies in the Hilbert space provide?

In Fig. 1(e), the Wigner functions of the cat qubit are experimentally obtained. Should they then not have proper axes?

In Fig. 1(e), why is the logical 0 state “approximately” equal to the coherent state α, while all other logical states (1, + & -) are “proportional to” linear combinations of coherent states α and -α?

In the caption for Fig. 1(e), “B1,· · ·, B5 (green) and “A1,· · ·, A4 (orange)” these ellipses ought to be lowered, not centered.

In the final paragraph of Section II, “The phase-flip times correspond to effective storage lifetimes under two-photon dissipation, T_{1,eff}, in the range 57–68 μs.” To be clear, this is referring to the effective T_1 of the storage cavity under two-photon dissipation? The authors have not stated the relationship between the cat qubit phase-flip time and T_{1,eff}.

In the third paragraph of Section III, “..the states 0_A = g and 1_A = f” should these A’s be lowercase a’s? Uppercase is inconsistent with the notation in the previous paragraph, unless I am missing something.

In the first paragraph of Section IV, “initialization of the ancilla A_i, two CX gates between A_i and its adjacent data qubits S_i and S_{i+1}” these A’s and S’s should be in italics to be consistent with the caption of Fig. 1(a).

Same paragraph, can the authors provide further detail on what is meant by the syndrome measurement cycle duration being “conservatively chosen’?

Second paragraph of Section IV, MWPM is used for decoding. There are excellent reasons for choosing MWPM for decoding of this code, but they are not stated here.

Section IV, in two places the logical phase flip probability per cycle is described as being equal to p_Z + p_Y, which is overly simplistic, based on some assumptions, and probably not what the authors mean anyway. If the noise was guaranteed to be stochastic Pauli errors, then one could define p_Z and p_Y (not done) such that this equation is true, but will that be the case here? No justification is given. There is a similar issue in the next section with the logical bit flip probability per cycle being equal to p_X + p_Y, and finally with a total logical error rate per cycle.

In the second-last paragraph of Section IV, “As the photon number increases, we see the expected increase in the logical error probability as the likelihood of a higher-order error not being caught by the error correction increases.” Does higher-order here refer to multiple phase-flips, or does higher-order refer to higher-degree polynomials of the photon creation/annihilation operator?

It would be preferable if Fig 5a and 5b could both use the same range for the y-axis, for easier direct comparison.

Referees' comments:

Referee #1 (Remarks to the Author):
The authors demonstrate quantum error correction (QEC) on a chain of 5 cat-qubits. Bit-flips are corrected by two-photon dissipation at the level of each cat-qubit. Phase-flips are corrected by an outer repetition code that measures the stabilizers XX for neighboring qubits. The entire QEC cycle is fast enough and introduces sufficiently low errors, that the logical phase-flip errors diminish when scaling the code from distance 3 to distance 5.

In my opinion this experiment is a landmark in the field of QEC.

Errors are the main roadblock towards the emergence of useful quantum machines. QEC provides a solution, in principle. However, currently investigated QEC architectures require daunting hardware overheads (e.g. surface code of transmons). There is a pressing need for new ideas that reduce this hardware overhead. Cat-qubits are one path forward (amongst many other promising ideas). The outstanding challenge is that although the entire QEC architecture is expected to be composed of a smaller number of qubits, and therefore easier to operate, each qubit is much more complicated than a transmon. There is therefore a compromise to find between (i) the complexity of the building block (transmon vs. cat-qubit ?) and the (ii) the complexity of the QEC architecture (surface code vs. repetition code ?). The experiment by Putterman et. al. shows that they are capable of mastering the fabrication and control of arrays of cat-qubits to a level where QEC actually works. Although these exotic dynamical qubits are still in their infancy, this rapid pace of progress is extremely encouraging to overcome the huge challenges lying ahead of universal fault-tolerant quantum computation.

The authors have made several bold and risky technological advances that have paid off.
(i) The matched chi ge - ef transmons, for fixed frequency transmons.
(ii) The low Kerr buffer modes by compensating the Kerr due to the cos(2phi) with the Kerr due to the junction array.
(iii) The flip-chip architecture with tantalum cat-qubit resonators
(iv) The tunable couplers between the transmon and cat-qubit modes.
Getting all these ingredients to work together on a chip of 5 cat-qubits, their 5 buffer modes and 4 ancillary transmons, is a true technological "tour de force".

I particularly appreciate that the authors are explicit on technical details like fabrication, fridge wiring and crosstalk cancellation. The peer-reviewing process in scientific journals is a gem to be protected. By supplying detailed information, referees (as well as the entire scientific community) may check the consistency of the results, challenge the conclusions and go as far as reproducing the experiment when necessary. In my opinion, it is important that all contributors to scientific journals (including companies) abide by editorial rules, and do not compromise the rigor of the reviewing process.

Before delving into more detailed and technical questions and comments, let me say upfront that I recommend this paper for publication in Nature. It is well written and the figures are of highly quality.

Prof. Zaki Leghtas.

****

We thank the reviewer for the detailed review of the paper and comments.

Detailed comments and questions:

1. In the abstract this sentence is misleading "We realize a noise-biased CX gate which ensures bit-flip error suppression is maintained during error correction". This may come across like a CX between two cat-qubits. Instead, what is performed is a CX between a cat-qubit and a transmon.

We added in a qualifier that this is a noise-biased gate between a transmon and a cat qubit. We note that we intentionally used the terminology “noise-biased” instead of “bias-preserving” to avoid confusion with gates such as the cat-cat CNOT gate which in

principle realize an unbounded suppression of the bit-flip error which we agree is an important distinction. Specifically the related sentence in the abstract now reads

> The logical bit-flip error is suppressed with increasing cat-qubit mean photon number, enabled by our realization of a **cat-transmon** noise-biased CX gate.

2. The authors mention multiple times the "threshold" of their code. They are careful in saying it is the "phase-flip correcting repetition code". For completeness they should however add a sentence in the text about the fact that overall (including bit-flips and phase-flips) this cat-qubit architecture does not possess a threshold. This is discussed in [Phys. Rev. A 103, 042413]. Although this may not be of practical significance, it is a conceptual fact that is worth mentioning.

We agree that this is an important subtlety when it comes to quantum error correction with biased noise qubits in a repetition code since for a given bias there will always be a lower bound on the logical bit-flip error rate set by the physical bit-flip error rates. We added a discussion of this to the total logical error rate section of the paper and referenced the suggested paper there as well. The paragraph we added to discuss this is shown below

> Importantly, we emphasize that a repetition code of biased noise qubits does not possess a proper asymptotic threshold because physical bit-flip errors are uncorrectable and place a lower bound on the achievable logical error~\cite{Guillaud2021_error}. Still, at large enough noise bias, the logical error can be reduced by increasing code distance before encountering this limit. In contrast, without bias, overall logical error would inevitably increase with code distance. The large noise bias in our device is apparent from the observations that we achieve comparable logical error as we increase distance, and that the distance-5 section outperforms the distance-3 sections for $l\alphal^2\geq 1.5$.

3. Minimizing the Kerr of the storage mode is crucial since the two-photon dissipation is pulsed (it is in fact off for a large fraction of the QEC cycle). This Kerr is partly induced by the Kerr of the buffer mode, which therefore needs to be minimized. The authors describe in a separate paper [ref 25] that they minimize the buffer Kerr by balancing the Kerr induced by the buffer junction array (negative) and the "cos(2phi)" contribution from the stray inductances in series with the small junctions of the ATS (positive). Could they explain how they fight against the Kerr induced by junction asymmetry? This is random, and difficult to measure at room temperature.

It is right that the side junction energy variation is a source of asymmetry and is random. While we do observe variation in the side junction energies away from the perfectly symmetric case, we nonetheless find that the variation is sufficiently tolerable that the stabilization works well on all of the unit cells. More quantitatively, the side junction energies on all of the unit cells are found to have an average asymmetry of ~1%. The worst unit cell has around 4% asymmetry but the storage self-Kerr nonlinearities of ~3 kHz (from numerical model of device) are still tolerable, though of course less would be better. We added a mention of this to the section of the supplementary material where we discuss the buffer mode (“buffer mode for implementing two photon dissipation”). Specifically it now reads

> As a result, the buffer-induced storage self Kerr and the storage-buffer cross Kerr are predicted to be small in most unit cells of our devices (i.e., $|K_{s}/2\pi| < 2~\text{kHz}$ and $|\chi_{sb}/2\pi| < 120~\text{kHz}$) compared to the 3WM interaction strength $g_{2}/2\pi$ which range from $320~\text{kHz}$ to $420~\text{kHz}$. In one of the unit cells which exhibits the worst bit-flip performance (i.e., S4), the buffer-induced nonlinearities are predicted to be large ($K_{s}/2\pi = 2.7~\text{kHz}$ and $\chi_{sb}/2\pi = 390~\text{kHz}$ compared to $g_{2}/2\pi = 350~\text{kHz}$) due to mis-targeting of the buffer junction energies. **The mis-targeting is specifically an asymmetry in the junction energies of the side junctions of the ATS.**

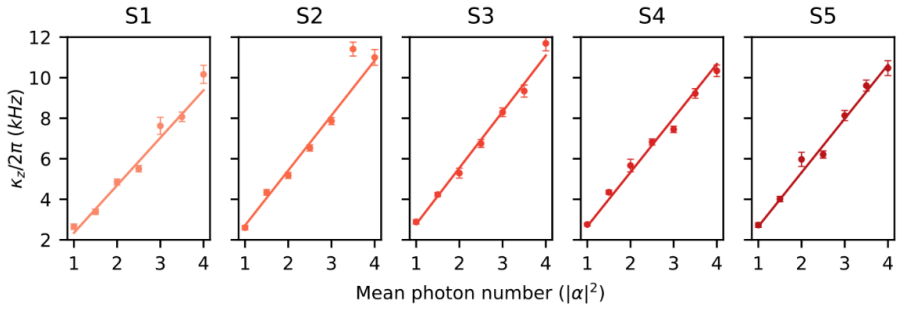
4. In Fig 1b some transmission lines cross. How was this achieved? Are there suspended bridges on the chip?

The chip uses 3D integration which allows us to put control lines on both the top and bottom chip. Lines that are routed on separate chips can cross without issue. We added a comment about this to the supplementary information in the section “Device fabrication” as follows

> The second, $20 \times 20~\text{mm}^2$ tantalum-based die is patterned to form storage resonators and other linear elements such as the readout resonators, Purcell filters for the readout resonators, and filters for the buffers. **Control lines are present on both the bottom tantalum-based die and the top aluminum-based die.**

5. In Fig 1f, I would suggest plotting 1/Tphaseflip (instead of Tphaseflip) in order to visualize the linear scaling as a function of alpha^2. The y-ticks could be in "us^(-1)" in order to immediately read-off Tphaseflip.

We agree that for a technical audience having the plot show $1/T_{phase-flip}$ is preferred to show the scaling is linear. We decided to show the lifetime since for a broad audience it would make it more immediate to see that the bit-flip times increase with photon number while the phase-flip times decrease with photon number. However, we added a plot of the phase-flip rates vs. $|\alpha|^2$ in the supplementary material in the section "Two photon dissipation calibration" to show the linear scaling for a more technical audience. We reproduce the plot below for your convenience.



6. In Fig 1e, there are missing ticks on the Wigners, and a missing colorbar.

We added in the missing color bar and ticks on the Wigner functions.

7. Combining the chi matching approach with two-photon dissipation where the latter corrects mis-rotations is a brilliant idea. This leverages the power of two-photon dissipation: it "cools" the system down to the cat-qubit quantum manifold. This solves the problem of leakage and error accumulation.

We thank the referee for this observation, and indeed the combination of chi-matching with two-photon dissipation is a critical part of this scheme.

8. Fig 2b: Is the Wigner plot post-selected for instances where the transmon was measured in g 10 times? i.e. is the smearing of the coherent state only due to Kerr or also to chi rotations due to the transmon jumping to e?

This data is from repeated application of the sequence of CX between storage-ancilla, ancilla reset, and ancilla readout. There is no post-selection applied. The dominant source of the smearing is more likely the Kerr nonlinearities, but in this case, the main point is that regardless of the error mechanisms the dissipative confinement is very helpful to prevent errors from accumulating over time.

9. Fig 2c. inset. Again, I would prefer the inverse of phase-flip time vs alpha^2 to verify the linear scaling. What is the solid line?

To keep this consistent with fig 1. we have left it as is. The solid line is a fit to the functional form $\kappa_z = \kappa_{1,eff}|\alpha|^2$ (we added mention of this to the text). In the supplementary material, fig 12, we show the linear fits as well for reference for a more technical audience.

10. Fig 2c. orange curve. Have you tried to post-select on having measured the transmon in g all the time? I would expect that the orange curve should then climb up to meet the black curve. The number of instances where you measure g at every shot over 10s of ms might be very low, in which case you could for e.g. progressively remove instances where you measured the transmon in e (or f) and see that depending on how many trajectories you throw away the orange meets the black.

We performed related post-selection experiments in the context of comparing the CX error rates with the transmon in the state |f>

to the CX error rates with the transmon in the state |g>. In the supplementary information in the section titled “Ancilla T1” we perform an experiment where we initialize the ancilla to |f> and post select against any experiments where the ancilla decays to |g>. Along similar lines to what is suggested we find that once we post select, initializing into |f> yields a performance similar to initializing into |g>. This indicates that indeed decay to |g> is the dominant source of error for the CX gate with the ancilla in state |f>. We haven’t done this post-selection style of experiment for the specific case of the ancilla in |g> which we leave to future work. We note that in doing such an experiment it may be important to post select on both the ancilla and the coupler being in the ground state since both may be subject to heating.

11. Fig 2a,b. The colorbars are missing.

We added the color bar to this figure (same used for a and b as mentioned in the caption).

12. Is CX^2 simply waiting twice as long as CX during the dispersive interaction of the transmon and storage?

At a high level it can be thought of that way, though it is slightly different because of the transients of the flux pulse. Specifically, the flux pulses we use are shaped at the beginning and end and during this shaped period the phase accumulation rate is slightly lower. As a result to achieve a full 2pi rotation the CX^2 will have a shorter length than just twice that of the CX. We added a citation to the supplement and also a mention that the CX^2 is a single pulse as shown below.

> We quantify the performance of our CX gate by repeatedly applying the pulse sequence of \cref{fig:architecture}(a) except with the single CX replaced by a single pulse that is the equivalent of two CX gates (a CX^{2} gate)~\cite{supplement}. This ensures that, similar to a stabilizer measurement, cat bit-flip times are first-order insensitive to ancilla state preparation errors.

13. Fig 3b: why are the first and last points above the average?

In the bulk a detection location is a comparison of syndrome measurements in subsequent rounds. At the beginning of the experiment the first stabilizer syndrome measurement is instead compared to the initial parities of the individual cat qubits that the stabilizer measures. For example, if the initial states of the cat qubits in the stabilizer were |+> and |->, the first stabilizer measurement would be compared to the fact that the initial joint parity was odd. In our case the state preparation of the storages to cat states is done with measurement. This means the first detection corresponds to a comparison of three measurements instead of two and thus we observe a higher error probability. The same is true for the final round where we compare to the final measurements of both cat qubits in the stabilizer. We added a comment on this to the supplementary material in the section “Decoding without erasure information” which we reproduce below.

> At the beginning and end of the repetition code experiment the syndromes are instead compared to the individual parities of the cat qubits that make up the stabilizer. This results in the temporal boundaries having different probabilities of having detectors triggered.

14. Fig 3d: how do you measure pmeas?

p_{meas} is extracted from the probabilities of the timelike edges of the error correction graph (which represent the effective syndrome measurement error probability). The edge probabilities are computed by looking at correlations between detection probabilities in subsequent rounds. In one case we post-select against all detection events that involve an erasure before computing the correlations. We detail this more in the supplementary information in the sections “Decoding without erasure information” and “Decoding with erasure information”. We added a brief note on this to the caption.

> (d) Effective **syndrome** measurement error **extracted from the QEC graph** before and after accounting for the erasure for $|\alpha|^2=1$.

15. Fig 3f: the meaning of the "faded points" is unclear to me. Also, they crowd the plot.

For small alpha $\lesssim 1$, the cat states have substantially different average photon number in $|+\rangle$ and $|-\rangle$ and therefore different phase flip rates. This means that the various product states of storage modes can have different lifetimes. The faded points serve

to indicate the spread caused by this effect and show the outcome from fitting the logical error rate binned by the number of excitations in the initial repetition code state. We discuss this more in the supplementary information in the section “Distribution of initial states in logical X lifetime experiments”. While we do appreciate this crowds the plot, we think this subtlety has important implications on interpreting the data and what range of photon numbers we consider (mentioned by another reviewer), and thus merits being included.

16. You mention "we carefully tune the syndrome measurements to avoid parasitic effects that can induce cat-qubit bit-flip errors, including [...] spurious two-level systems (TLSs) in the ancillas". How do you avoid TLS? Is just by thermal cycling until you are lucky enough that no TLS couples to any transmon?

Yes, thermal cycling can rearrange the ancilla TLS distribution and we used that if ancillas (or other modes) had TLS in particularly problematic locations. One can also change the coupler position which through the coupling to the ancilla can change the interaction position (i.e. where the gate happens) of the ancilla. Ultimately this optimization was not used for the data presented in the paper. Additionally, there are other considerations such as ensuring that the readout does not shift the ancilla into a TLS due to the dispersive shift. Going forward one could move to flux tunable ancillas to more robustly avoid TLS. The paragraph in question was removed from main text for length reasons but details about TLS are present in the supplement.

17. About Sec V: since you have access to single shot Z measurements, and that in principle this measurement is QND, have you measured single trajectories for Z? If so, have you looked for correlations between Z trajectories of different cats?

This is not something that we have done. If we understand correctly, the suggestion is to measure the cat qubits in the Z basis during every cycle of the quantum error correction experiment and study how the Z observable of an individual cat qubit evolves over time. This isn't something we've done since the individual Z basis measurements don’t commute with the stabilizer measurements for error correction. However, this would be indeed interesting to look at in future works, e.g., in the context of simultaneous stabilization of multiple cat qubits during idling.

18. Section VI: you write that for the distance 5, the minimal measured eL = 1.65% ± 0.03%. For distance-3 sections you find eL = 1.83% ± 0.03% and eL = 1.67% ± 0.04%. I am assuming the eL = 1.75% quoted in the abstract is the average the two latter numbers. This is a bit misleading, and contrasts with the otherwise very honest and clear claims of the paper. The results of this paper are very impressive, so there is no need to make a controversial claim in the abstract (comparing d=5 to the average of the two d=3 codes, knowing that one of the d=3 codes had the same eL than the d=5). For the abstract, I encourage the authors to highlight the results of Fig3f, where we see that the d=5 outperforms both d=3 codes.

In computing the average of the distance 3 sections we are following the procedure used elsewhere in the literature (see Nature 614, 676–681 (2023)) to get a representative number for a given distance. It has never been out intention in the abstract (and in the rest of the paper) to make any below threshold claim about distance-3 vs. distance-5 overall logical error probability due to the lack of an explicit threshold in a repetition code of biased noise qubits. We mainly wanted to show that in spite of the increased number of fault locations the distance-5 code performs comparably to the distance-3 code which is a feat in and of itself. We clarified the wording in the abstract and also added the average error rate into the main text to make it clear how it was computed. Furthermore, we think the explicit discussion added about the lack of a threshold will make this even more clear. We thank the reviewer for pointing out how this could be misinterpreted. The relevant section of the abstract is shown below

> The minimum measured logical error per cycle is on average $1.75(2)\%$ for the distance-3 code sections, and $1.65(3)\%$ for the distance-5 code. The comparable performance, despite the increased number of fault locations of the distance-5 code, is enabled by the high degree of noise bias preserved during error correction.

The main text was updated as follows to show the average error rate

> This is comparable to the best observed performance for the distance-3 sections which are $\epsilon_{L}=1.83\% \pm 0.03 \%$ and $\epsilon_{L}=1.67\% \pm 0.04 \%$ at $|\alpha|^2=1$ **(average $\epsilon_{L}=1.75\% \pm 0.02 \%$)**

19. The last paragraphs in the conclusion are very instructive. The authors talk about the "performance floor" due to the transmon. Could they be more explicit on the numbers? What if you could make a 1 ms transmon (or Fluxonium), what would that noise floor be? Since you are only sensitive to the double decays during the CX, the numbers might be quite good. It is worth mentioning them, because this strategy could go a long way.

We agree there are significant opportunities with this architecture by extending transmon T1s. We added a number for the noise floor with longer transmon T1s as suggested. We note that at some point the heating of the transmon will start to become the limiting factor. We would also like to point out that in the appendix we discuss more detail about the T1 error model for the ancilla. In the current gate, we find that the CX gate error is higher than predicted from the independently measured T1 values. Understanding and making the CX gate induced bit-flip errors fully ancilla coherence limited will also allow for lowering the performance floor.

> The use of ancillary transmons for syndrome measurements is critical to our experiment, enabling noise-biased CX gates without undesired control errors, but comes along with ancilla induced bit-flip errors. Current logical performance is not limited by these errors, and improved ancilla lifetimes of 1ms would ideally lower their probability to $\sim 10^{-6}$ per CX~\cite{supplement}).

20. There are multiple loops on this chip, could the authors say something about flux drifts: how they minimize them (shielding, filtering etc) and how they actively compensate them (recalibrations?). How large are these drifts (in units such as mPhi0 per hour)?

We haven't quantified the flux drifts in the system as we generally observe that the system fluxes are relatively stable over time. We have multiple levels of magnetic shielding in our system comprised of mu-metal shields at room temperature and mu-metal and superconducting shields at the MXC stage of the fridge. The recalibrations performed for repetition code experiments are detailed in the supplementary information in the section "Repetition code characterization procedure". In essence we perform Ramsey (and related measurements) to capture any small changes in mode frequencies. Note that these frequency changes could also be caused by off resonant spurious modes moving and would not necessarily be due to flux drift. We also point out that with the components in their idle positions the buffer and couplers – the flux tunable elements in the system – are at first order flux insensitive positions. We added a mention of the magnetic shielding configuration into the supplementary information shown below.

> The package the chip is mounted in is shielded by mu metal and superconducting shields at the bottom stage of the fridge in addition to mu metal shielding at room temperature.

Referee #2 (Remarks to the Author):
The authors report on a quantum logical memory experiment, realizing a distance-5 repetition code of bosonic cat qubits, on a superconducting circuit architecture. The cat qubits utilize higher levels of harmonic oscillators to encode the quantum information in (a superposition of) coherent states stabilized by two-photon dissipation, with a large noise bias (bit-flip time much larger than phase-flip time). Capitalizing on this noise bias, the authors can afford to correct only one type of error, therefore requiring only a one-dimensional repetition code, rather than a two-dimensional quantum error correction (QEC) code, therefore justifying the "hardware-efficient" qualifier.

This is a novel result: cat qubits were demonstrated on their own, and planar QEC codes such as the surface code as well, but the repetition code of noise-biased bosonic qubits is a first (to my knowledge). It is also not a trivial result, because of the requirement to preserve the noise bias in the syndrome measurements. The authors cleverly address this using a g-f encoding of the ancilla qubits, matching the dispersive shifts, and detecting erasures when the f-state decays to e.
I also find the experiment to be timely, with many experimental quantum logical memory results appearing in the past few years (Ref 28-35), at or below threshold.

Generally, the data is of high quality and supports most of the claims in the manuscript (besides one, see below). The figures are well-presented, with error bars, and reliable trends.

We thank the reviewer for the detailed review of the manuscript and the helpful suggestions.

My two main suggestions for improvements are the following:

1. The claim to be below threshold is a bit tricky, as the repetition code distance indeed improves the phase-flip time (Fig 3f), but decreases the bit-flip time (Fig 4b). What ultimately matters is the logical memory error rate. There the description is suddenly

vague, such as when the abstract reads "demonstrating the effectiveness of bit-flip error suppression throughout the error correction cycle." (I am honestly not sure I parse this sentence as the authors intended.)

We agree that claiming the repetition code is below threshold is not a fair claim which is why we were careful to only say that the "*phase-flip correcting repetition code operates below threshold".* In general, for a repetition code of biased noise the notion of a threshold is more complex since for a given bias there will always be a point where the error rate starts to increase with distance. We have added a discussion into the paper in the total logical error rate section to make sure this is addressed.

> Importantly, we emphasize that a repetition code of biased noise qubits does not possess a proper asymptotic threshold because physical bit-flip errors are uncorrectable and place a lower bound on the achievable logical error~\cite{Guillaud2021_error}. Still, at large enough noise bias, the logical error can be reduced by increasing code distance before encountering this limit. In contrast, without bias, overall logical error would inevitably increase with code distance. The large noise bias in our device is apparent from the observations that we achieve comparable logical error as we increase distance, and that the distance-5 section outperforms the distance-3 sections for $|\alpha|^2\geq 1.5$.

The abstract was also updated to make the discussion more clear

> The minimum measured logical error per cycle is on average $1.75(2)\%$ for the distance-3 code sections, and $1.65(3)\%$ for the distance-5 code. The comparable performance, despite the increased number of fault locations of the distance-5 code, is enabled by the high degree of noise bias preserved during error correction.

I believe it would make sense to bring some clarity and be transparent about:
a) Why is the minimal value of |alpha|^2 (=1) also the one that represents the best results in the distance-3 codes? Are the authors sure that going below doesn't improve the performance even further? It seems odd in Fig 5a to have the optimum as the extremum of the explored range.

At low photon number there is an asymmetry between the phase-flip error rate of the even and odd cat states. Intuitively this is clear in the limit of $|\alpha|^2 = 0$ where the even cat is the 0 photon Fock state and the odd cat is the 1 photon Fock state. Thus as the photon number is lowered this leads to a spread in the error rate for initial states with a different number of even cat states. In addition, this can lead to an artificially lower error rate at low photon number (benefiting the distance-3 code more than the distance-5 code). As can be seen by the faded points, which represent the logical error when grouped by the number of even cats in the initial state, the dependence on the initial state already causes a large spread in error rates at $|\alpha|^2 = 1$. Going to lower error rates will further increase this spread (exponentially) and make the logical lifetime less clearly defined. We go into more detail on this in the supplementary material in the section titled "Distribution of initial states in logical X lifetime experiments".

We also note that the interpolation of the overall logical error rate shows that the minimum distance-3 performance occurs between 1 and 1.5 photons. This interpolation is based on an interpolation of both the logical bit-flip and logical phase-flip error probabilities, both of which are monotonic at low photon number, as a function of $|\alpha|^2$.

b) What is the logical error suppression factor when increasing the hardware code size (overall seems to be possibly 1.75/1.65=1.06 with a large error bar)?

For the reasons mentioned earlier about there not being an explicit notion of a threshold for a biased noise repetition code we choose not to quote a logical suppression factor as this could be misleading to the readers.

c) is the concatenated-cat code beyond any break-even (if the information was instead encoded in a Fock state of one of the harmonic storage mode)? Seems to be just at or around break-even.

Note: I don't believe it would reduce the impact of the paper, rather provide a clean and easily understandable status.

We added in a section of the supplementary material ("Breakeven comparison") where we perform a breakeven comparison to the 0/1 Fock encoding of the storage modes. Indeed the performance is close to but about ~20% below the breakeven point of the average storage mode 0/1 Fock encoding.

2. The text clarity could be improved at various places (often refers to statements without enough details to follow, yet at times wondering if really required, or could be made much shorter). It gives an overall feeling of being rushed during writing.
Some examples:
a) The conclusion has 4 paragraphs, which read:
(i) summary of the work (ok),
(ii) what is limiting and could be done better, *as demonstrated in Manuscript in preparation* (!! is ithttps://arxiv.org/abs/2409.17556?),
(iii) one could then tile a 2D code anyways instead of 1D, see analysis in *Manuscript in preparation* (!!),
(iv) what could happen hypothetically (in too much details) if the problem of implementing bias-preserving gates between cats is solved.
Paragraphs (ii)-(iv) could be combined in a single much shorted one, that provides the same ideas with more punch and less reliance on (uncertain) future works.

We restructured the conclusion to simplify some of the more technical details. We also added in the full reference for the future works which have been released in the past weeks. We kept the paragraphs to separate out the ideas since other reviewers commented a different point of view on it giving a nice sense of future directions.

b) "Moreover, the buffer-mode parameters are carefully chosen to minimize other parasitic buffer-induced nonlinearities on the storage mode."
What parameters? What other parasitic nonlinearities? This is a filler sentence that either needs expansion, or be removed completely and just refer to the suppl. mat.

We were attempting to highlight some of the innovations that enabled this experiment but we agree that without more detail it is not constructive. We removed this detail from the paper and just referred to our other paper and the supplementary information. It now reads as follows

> To realize the two-photon dissipation, we nonlinearly couple each storage mode to a lossy buffer mode (green) which is implemented using a version of the asymmetrically-threaded SQUID element~\cite{Lescanne2020,singlecat2024}. Crucially our buffer mode implementation ensures that the storage-mode lifetime and linearity are not degraded by the coupling to the lossy and nonlinear buffer~\cite{singlecat2024, supplement}.

c) "Only higher-order or suppressed ancilla error mechanisms, such as two sequential decay events or heating will cause bit-flip errors on the data cat qubit."
Which mechanisms are considered higher-order or a suppressed? What is the difference? Aren't they suppressed because they are higher-order?

It's more semantic than anything, but we were being careful because double decay is second order in the ancilla T1 whereas heating is first order in the ancilla T1 but suppressed by an additional Boltzmann factor. We simplified the wording based on your comment.

> Only subleading ancilla error mechanisms such as two sequential decay events and heating will cause bit-flip errors on the data cat qubit.

d) "The small degradation relative to the reference performance at |alpha|2 >~ 3 is due to heating events in the ancilla or coupler during the CX^2 gate."
How high is heating? Is a factor ~2 difference at |alpha|^2=4 really a "small difference"?

The degradation in the CX bit-flip time from having ancilla population in the |f> state (orange relative to blue), which can then lead to bit-flip errors if the ancilla decays to |g>, is much larger than the bit-flip error due to heating (orange relative to black). The coupler/ancilla heating rate is on a timescale longer than a millisecond. Following the comment we removed the word “small” since it was unnecessary to the discussion and could confuse the reader. It now reads as follows

> The degradation relative to the reference performance at $|\alpha|^2\gtrsim 3$ is due to ancilla or coupler heating during the CX^{2} gate.

e) "we show a control sequence involving a CX gate similar to the one used for an error correction syndrome measurement."
Is it not equal? What makes it a "similar" gate rather than the same?

The most notable difference is that for error correction syndrome measurement we perform two CX gates. Specifically there is one CX gate to each of the storage modes adjacent to the ancilla transmon. In contrast for the case of the CX characterization we just perform a gate to one storage mode to isolate each interaction's performance. The specific gate we perform is actually a CX^2 gate (a 2pi rotation) which is important to ensure that analogously to a stabilizer measurement the storage mode bit-flip probability is not sensitive to state preparation error. There are some other more minor differences, for example slightly different cycle lengths. This is discussed in more detail in the supplementary material section "CX (and CX^2) gate calibration procedure". We added a citation to the supplement section as shown below

> We quantify the performance of our CX gate by repeatedly applying the pulse sequence of \cref{fig:architecture}(a) except with the single CX replaced by a single pulse that is the equivalent of two CX gates (a CX^{2} gate)~\cite{supplement}.

f) "Notably, all these imperfections can be corrected with high probability when the two-photon dissipation is turned back on after the CX gate"
How high are we talking?

This depends on the nature of the imperfection. One example of an imperfection is misrotation of the storage mode state which can for instance be induced by imprecise χ-matching. In the supplementary material section titled "CX χ-matching tolerance" we show data for the bit-flip probability as a function of the amount of χ-matching when applying CX^2 gate. As can be seen for a mismatch of up to ~10% the bit-flip probabilities are still strongly suppressed (increased by $\lesssim 10^{-3}$ compared to a baseline of $\sim 2 * 10^{-3}$).

g) "Each syndrome measurement cycle has a conservatively chosen duration"
Can you be quantitative on "conservatively"? How much buffer time is planned?

We discuss the exact pulse sequence in more detail in the supplementary material sections titled "Repetition code X basis pulse sequences" and "Repetition code Z basis pulse sequences". In those section one can also find the full pulse sequence used in the experiment. Roughly speaking with some optimization the cycle time could be shorted to $\sim 2\ \mu s$ from $2.8\ \mu s$. Further improvements can be achieved by achieving more uniform gate durations. We added a citation to the supplementary material as shown below.

> Each syndrome measurement cycle has a conservatively chosen duration of 2.8~μs~\cite{supplement}.

h) "we carefully tune the syndrome measurements to avoid parasitic effects that can induce cat-qubit bit-flip errors, including [...] spurious two-level systems (TLSs) in the ancillas"
I don't see any tuning of syndromes for this. The suppl. mat. only refers to "temperature cycle the dilution fridge".

That is correct that we use temperature cycling to tune TLS if they are in particularly problematic locations in our modes. We also have some flexibility to adjust the ancilla frequency in the interaction position by varying the coupler position but in the final configuration used for the experiment the couplers are at their minimum frequency positions. In the future, one could use flux tunable ancillas to simplify the TLS avoidance (mentioned in supplementary material). The details about the repetition code calibration are still discussed in the supplementary material.

i) Fig1a and b would benefit from labeling Si, Ai, ...

We added in the labeling of the 5 storage modes used in the experiment to fig 1b. We agree this makes seeing the mapping to the high level schematic of the device mode clear. We didn't add the ancilla label since the figure started to get busy and we think the recommended storage label combined with the zoom ins give enough information.

Referee #3 (Remarks to the Author):
The submission "Hardware-efficient quantum error correction using concatenated bosonic qubits" demonstrates experimentally

the use of biased-noise qubits in a quantum error-correcting memory. The physical qubits are bosonic cat qubits, which are notable for their extreme noise bias. It is natural to examine the use of a biased cat qubit in a quantum repetition code, where phase-flips are corrected at the logical level, while bit-flips are suppressed at the physical level.

This submission presents an impressive and timely research result in quantum error correction. Its impact and level of scientific interest will be high, given recent results in surface code quantum error-correcting memories, with which these present results are very complementary. The key technical innovation in this paper is the demonstration of a noise-biased CNOT gate, which underpins the scheme, and this is combined with an impressive integration of a number of previously-demonstrated components and techniques into a system that performs in a regime relevant to fault-tolerant quantum computing. We have not identified any fundamental flaws that would prevent publication. For these reasons, we believe the paper is clearly of the standard that warrants publication in Nature.

The paper is well-written, and the flow of results from the level of physical qubits to the level of logical memory makes the paper easy to read for a broad audience. This comes at the cost of specificity, but this is unavoidable given the length constraints.

We thank the reviewer for the detailed review of the manuscript and helpful comments.

We have provided a number of comments and questions for the author's consideration.

Comments:

In the introduction, "A complementary QEC paradigm is to use a layered approach to noise protection, in which one starts from a qubit encoding that natively protects against different noise channels and suppresses errors." While recognising that this statement is somewhat 'setting the scene', it is not entirely accurate that an approach of concatenating a bosonic code with a repetition code sits in a different paradigm than surface-code-based QEC. We'd encourage the authors to state more directly the differences between this approach and others.

We have updated the wording to remove "complementary paradigm" and make the emphasis more explicitly on using an encoded qubit instead of a simple physical qubit. This was the original intention but the new wording should make this more clear.

> Alternatively, one can use a layered approach to noise protection by starting from an encoded qubit that natively suppresses errors.

"One example is bosonic qubits, where qubit states are encoded in the infinite-dimensional Hilbert space of a bosonic mode (a quantum harmonic oscillator), and the extra dimensionality of the mode provides redundancies that can be utilized to perform bosonic QEC." We appreciate the need for brevity, could the authors elaborate on what exactly the redundancies in the Hilbert space provide?

We improved the wording for this paragraph since the previous mention of redundancy was confusing. The new version (shown below) focuses more on the large oscillator Hilbert space.

> An example is bosonic qubits, where qubit states are encoded in the infinite-dimensional Hilbert space of a bosonic mode (a quantum harmonic oscillator) using bosonic QEC~\cite{Cochrane1999Macroscopically, Gottesman2001, Jeong2002Efficient}. In bosonic QEC, the large oscillator Hilbert space is exploited to suppress errors. Experiments demonstrating this at the single bosonic mode level have been performed using cat codes

In Fig. 1(e), the Wigner functions of the cat qubit are experimentally obtained. Should they then not have proper axes?

We have added in the axis in these plots.

In Fig. 1(e), why is the logical 0 state "approximately" equal to the coherent state α, while all other logical states (1, + & -) are "proportional to" linear combinations of coherent states α and -α?

This is a subtle point about cat qubit state normalization. The sum even and add cat states are directly proportional to the sum of $|\alpha\rangle$ and $|-\alpha\rangle$. Importantly the normalization factor is different between the even and odd cat states. Specifically the normalization is given by $N_{\pm}(\alpha) = \frac{1}{\sqrt{2(1\pm e^{-2\alpha^2}}}$. Since the normalization is not the same between even and odd cat states this results in $|0\rangle \propto |+\rangle + |-\rangle$ containing a contribution from $|-\alpha\rangle$ which is exponentially small in mean photon number. We have added further discussion of this to the supplementary material

> Thus, the computational basis states $|0\rangle = (|+\rangle + |-\rangle) / \sqrt{2}$ and $|1\rangle = (|+\rangle - |-\rangle) / \sqrt{2}$ are approximately given by the coherent states, i.e., $|0\rangle \simeq |+\alpha\rangle$ and $|1\rangle \simeq |{-}\alpha\rangle$ in the limit of $e^{-2|\alpha|^{2}} \ll 1$. **Note the reason it is only approximate is because of the difference in normalization between the even and odd cat states.**

In the caption for Fig. 1(e), “B1,· · ·, B5 (green) and “A1,· · ·, A4 (orange)” these ellipses ought to be lowered, not centered.

Thanks for catching this, we lowered the ellipses.

In the final paragraph of Section II, “The phase-flip times correspond to effective storage lifetimes under two-photon dissipation, T_{1,eff}, in the range 57–68 μs.” To be clear, this is referring to the effective T_1 of the storage cavity under two-photon dissipation? The authors have not stated the relationship between the cat qubit phase-flip time and T_{1,eff}.

We added in the definition of $T_{1,eff}$. It was an oversight that we did not originally have this. $T_{1,eff}$ represents the effective T_1 of the storage mode extracted from the phase-flip rate as a function of $|\alpha|^2$. The new discussion is shown below

> As expected, the phase-flip times degrade as $T_{1,\text{eff}}/|\alpha|^2$, where the effective storage lifetimes under two-photon dissipation, $T_{1,\text{eff}}$, are in the range $57-68$~μs.

In the third paragraph of Section III, “..the states 0_A = g and 1_A = f” should these A’s be lowercase a’s? Uppercase is inconsistent with the notation in the previous paragraph, unless I am missing something.

Thanks for catching this, we fixed the notation in section III to be consistent. (reproduced below)

> As in Refs.~\cite{Rosenblum2018,Reinhold2020, Ma2020} we encode the ancilla qubit into the states $|0_{a}\rangle = |g\rangle$ and $|1_{a}\rangle = |f\rangle$ and engineer an approximately ``χ-matched'' dispersive interaction between the ancilla and the storage mode.

In the first paragraph of Section IV, “initialization of the ancilla A_i, two CX gates between A_i and its adjacent data qubits S_i and S_{i+1}” these A’s and S’s should be in italics to be consistent with the caption of Fig. 1(a).

Thanks for catching this, we fixed the notation in Section IV to be consistent. (reproduced below)

> comprises initialization of the ancilla $\text{A}*{i}$, two CX gates between $*\text{A}*{i}$ and its adjacent data qubits $\text{S}*{i}$ and $*\text{S}*{i+1}$, and finally measurement and reset of the ancilla.

Same paragraph, can the authors provide further detail on what is meant by the syndrome measurement cycle duration being “conservatively chosen’?

In the supplementary information sections titled “Repetition code X basis pulse sequences” and “Repetition code Z basis pulse sequences” we provide more detail. We added a citation to the supplement so the reader can find this additional detail.

> Each syndrome measurement cycle has a conservatively chosen duration of 2.8~μs~\cite{supplement}.

Second paragraph of Section IV, MWPM is used for decoding. There are excellent reasons for choosing MWPM for decoding of this code, but they are not stated here.

We added some references to papers that discuss MWPM which can provide more discussion on the advantages of MWPM.

> After running an experiment with many error correction cycles we decode the syndrome measurements using minimum-weight

> perfect matching (MWPM)~\cite{pymatchingv2, Fowler2012}.

Section IV, in two places the logical phase flip probability per cycle is described as being equal to p_Z + p_Y, which is overly simplistic, based on some assumptions, and probably not what the authors mean anyway. If the noise was guaranteed to be stochastic Pauli errors, then one could define p_Z and p_Y (not done) such that this equation is true, but will that be the case here? No justification is given. There is a similar issue in the next section with the logical bit flip probability per cycle being equal to p_X + p_Y, and finally with a total logical error rate per cycle.

We discuss this error rate convention more in the supplementary information in the section titled “Basis and error rate convention”. In the main text we only include the expression in terms of the logical Pauli error probabilities for the overall logical error probability. When doing so we also put in parenthesis that this is assuming a Pauli error model. We agree this is an important qualifier since at low photon number (n <= 1) the asymmetry in the phase-flip error rate between different logical states may result in something which is not exactly a Pauli error model. The remaining location this is mentioned is shown below

> Combining together the logical bit-flip and phase-flip probabilities, we show in \cref{fig:total_lifetime}(a) the overall logical error per cycle \cite{Krinner2022,Acharya2023}, $\epsilon_{L}= (\epsilon_{L,\text{phase-flip}} + \epsilon_{L,\text{bit-flip}}) / 2$ **($(p_X + 2p_Y + p_Z)/ 2$ under Pauli noise)**, for the repetition cat codes.

In the second-last paragraph of Section IV, “As the photon number increases, we see the expected increase in the logical error probability as the likelihood of a higher-order error not being caught by the error correction increases.” Does higher-order here refer to multiple phase-flips, or does higher-order refer to higher-degree polynomials of the photon creation/annihilation operator?

Higher-order refers to the former. As the photon number increases, the probability of many phase-flip errors occurring and confusing the error correction decoder increases (specifically (d+1)/2 errors in the ideal case). We modified this statement as we agree the wording is confusing. Now we just make the statement that the logical phase-flip rate increases because the physical phase-flip rate increases. We leave the reader to draw the conclusion that this is because the probability of higher weight errors across the cat qubits are more likely as photon number increases. The new discussion is shown below

> As expected, as the photon number increases, the logical error probability increases because the cat qubit phase-flip rates increase.

It would be preferable if Fig 5a and 5b could both use the same range for the y-axis, for easier direct comparison.

We have changed the axis labeling to make it consistent between the two plots.