nature portfolio

Peer Review File

A fault-tolerant neutral-atom architecture for universal quantum computation

Corresponding Author: Professor Mikhail Lukin

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.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 manuscript by Bluvstein and Geim et. al. demonstrates several building blocks towards a practical logical quantum processor with fault-tolerant universal operations. They first implement projection onto logical qubit subspace via stabilizer measurement, successfully demonstrate logical error suppression and compare different logical entangling methods using surface codes. They then implement synthetic arbitrary-angle single-qubit rotations with the [[15, 1, 3]] code, and finally explore logical-level deep-circuit behavior using logical teleportation with [[7, 1, 3]] and [[16, 6, 4]] codes. Several hardware updates, including spin-to-position non-destructive readout and loading from a reservoir, are employed for better hardware performance and faster clock speed. The results are solid, revealing a promising roadmap towards universal fault-tolerant quantum computation using reconfigurable neutral atom arrays.

The authors declare that in this work they “implement all key elements of a universal, fault-tolerant quantum processing architecture”. Considering all the impressive results in this work, I would still find this claim slightly exaggerative, with reasons listed as follows:
a. General fault-tolerant stabilizer measurement is missing — the NZNZ pattern for surface code is FT, but would not work for codes such as [[7, 1, 3]] and [[15, 1, 3]] which are used in logical non-Cliffords and logical teleportation below. Shor/Steane/Flag type stabilizer measurement protocol is in general required.
b. Mid-circuit error correction after decoding is missing — The authors claim that +1-logical subspace is required for non-Clifford gates, but it is not demonstrated.
c. General non-Clifford operations are missing — The synthetic arbitrary-angle rotation using [[15, 1, 3]] code is based on the specific transversal universal gate set of the code (more specifically, transversal CZ available), and is not guaranteed to work with other codes.
d. The logical teleportation protocol for cleaning up physical-level errors is also based on transversal CZ gates.
All those are parts of my major concern regarding the results: How do all these building blocks come together to perform a FT logical quantum processor using codes with decent encoding rate/distance, but without built-in good transversal gate set (qLDPC for example, which may only have transversal logical Paulis and CNOTs)? While the claim is overall not hurting the solidity of the results, I would like to see the authors’ thoughts on the steps forward before I can fully recommend this work to be published in Nature.

Besides that, I have several more questions and suggestions for the authors to consider before resubmitting their work:
1. For below-threshold performance:
a. What is the source of that ~8% detector probability with post-selection (Fig. 2b and 2c)? An error budget would help with better understanding of the hardware performance.
b. I find the explanation of the four metrics in Fig. 2(b)(c) a bit ambiguous:
i. Are the atom loss information from a final readout of the data block, or from observing the flickering pattern of stabilizers?
ii. Are the four groups of detector loss data extracted from the same experimental dataset and the only difference is the data processing?
iii. If the answer to the previous problem is yes, then I don’t understand the difference between “bare” and “detect loss”: doing an entangling gate with an atom in |0> state is conceptually different from doing an entangling gate with nothing (atom loss).

It is confusing how to count loss as state |0> or not in the data processing under a pre-determined physical process.
iv. If “post.” is post-selecting on no atom loss, then where does the “post.” data with non-zero loss rate come from?
c. For ML decoder, do the authors have an estimation of how the time complexity scales with the code size?
2. General comment: For several sets of data analysis, the choices of acceptance fractions are vaguely clarified. Detailed explanations would be helpful.
3. For the non-destructive readout, it seems appropriate to cite other approaches in the community that have been used to demonstrate QND readout quite a while ago in some cases. For instance, stretched readout techniques with alkalis were demonstrated by Meschede and Saffman in 2017 and are used in large-scale alkali processors by Saffman and Pritchard, QND readout of optical clock qubits was realized and exploited by Endres and Kaufman in 2019, and QND readout of nuclear spin qubits was realized and exploited by Covey and Atom Computing in 2023. As such, the novelty of this capability here seems to be somewhat overstated, as only the Weiss approach that was essentially duplicated here is mentioned. That said, I appreciate the demonstrations of the significance of QND readout for deep circuits.
4. As a related point, the community would benefit from must faster QND readout. The readout time does not seem to be directly mentioned, but it appears to be 3~10 ms depending on the settings. The authors should comment on the prospects to bring this down to the ~10-100 μs level. This also goes for cooling and reset. Given that destructive measurement can be performed in ~10 μs, it may be appropriate for the authors to briefly present their vision for whether slow non-destructive measurement is better than fast destructive measurement, particularly with the advent of coherent reloading to obviate atom loss constraints.
5. In Figures 1c &d, presumably the “QEC” refers only to stabilizer measurement and not active correction via conditional operations. Although the passive correction or conversion due to the Zeno effect is very nice, the language surrounding this study is somewhat misleading. If the authors were to actually do syndrome extraction and correction in each “QEC” round, to what extent is it meaningful to “track the state of the system” to detect incoherent errors? Perhaps coherent errors would be corrected on each round, and then incoherent errors would be inferred based on comparing several rounds?
6. For transversal CNOT vs. lattice surgery, it is confusing that the transversal CNOT error rate does not change at all at various ancillary measurement errors (Fig. 3c). Is there no error correction (even in software) required after stabilizer measurements if the measurement is just for projecting to logical subspace/remove entropy? Is it some properties related to correlated decoder?
7. As a related point, I find it less compelling that the “data” is entirely from post-processing. What role is the experimental data actually playing here? Is the experimental data used shot-by-shot, or is it averaging over many realizations with the post-processing (injecting measurement errors) on averaged data? Why is this not just a theory result in terms of what error probability per logical operation we expect for a given ancilla measurement error? Or at least, why are there not theory lines on Fig. 3(c)? It seems like the role of real data is to set the offset (error probability per logical operations when there is no injected ancilla measurement errors), but the trend should presumably entirely match a simple theory.
8. As a related point, how would this story change if d rounds of measurements were used for lattice surgery instead of just one? It seems straightforward to implement this in post-processing and straightforward to have a theoretical expectation. To be clear, I am not suggesting that lattice surgery is better than transversal gates, but the presented studies seem somewhat contrived.
9. The issue of feedforward comes up again for the synthesis of T states. Since the feedforward operations are only being applied in software, true T states are presumably only generated probabilistically via post-processing. As such, the authors should comment on the “yield” and how it scales with code distance. They should also comment on whether such teleportation protocols are deterministic (when logical feedforward is applied) and how scalable this approach is to larger codes. It may also be appropriate to provide a fidelity for T state preparation. The note about ED fig. 9f and the CHSH inequality suggests that the fidelity is several 9’s, but the discussion in the main text is surprisingly qualitative.

Referee #2

(Remarks to the Author)
Report on manuscript Architectural mechanisms of a universal fault-tolerant quantum computer

In their work, the authors present an experimental demonstration of all essential components required to perform universal error-corrected fault-tolerant logical quantum computation on multiple logical qubits. This is enabled by a number of key improvements to the experimental architecture, as compared to previous work by earlier work by the team demonstrating a logical quantum processor. Specifically, the authors introduce a technical upgrade of their neutral atom quantum processor, transitioning from spin-to-loss to spin-to-position conversion. This enables non-destructive state readout and additional loss detection at the end of a computation. On this platform, the authors demonstrate repeated rounds of QEC on the rotated surface code of distance d= 3 and d= 5, operating with approximately a factor of 2 below threshold performance. They extract logical error rates for up to five rounds of quantum error correction (QEC) and achieve a similar error-suppression per round as shown in [1, 2]. Experiments are accompanied by numerical simulations, which allow the authors to estimate the relative error contributions of different physical error mechanisms. Building on their QEC protocols, they implement logical entangling gates by means of transversal CZ-gates interleaved with QEC. They find an optimal logical error per logical operation for three CZ-gates per QEC round, and conclude that the entropy introduced by the logical entangling gates is exactly balanced by the stabilizer measurements. They further realize logical entangling operations through lattice surgery, for the first time demonstrated in a neutral-atom processor, by means of joint logical ZZ-measurements on an initial state |++⟩ and achieve error rates comparable to recent works [3]. The authors then prepare distance-3 2D and 3D color codes in |+)L and observe robust angles under rotation about the Z-axis, which correspond to stabilizer revivals. Arbitrary single-qubit rotations are then synthesized on the 3D color code with transversally realized teleportation-based protocols, where the Eastin-Knill theorem is circumvented with (non-unitary) logical measurements. Based on these protocols, the authors inject layers of logical HT-gates, thus preparing many states on the Bloch sphere that cannot be reached with purely unitary transversal operations by any single QEC code, conceptually similar to earlier demonstrations with trapped ions [4].

Furthermore, the authors implement deep logical circuits comprising of up to 27 layers of transversal entangling operations on 32 instances of the distance-3 2D color code. They leverage teleportation protocols such that physical errors are removed while the logical information propagates through the circuit. For
the prepared logical 1D and 2D cluster states, they can preserve correlations substantially longer than for physical qubits. Lastly, the authors encode high-rate codes and, again, execute transversal entangling gates to prepare large cluster states and relialy teleport logical quantum states forward in a measurement-based QC fashion.

The key milestones include repeated rounds of quantum error correction (QEC), logical entangling gates, a protocol for arbitrary single-qubit logical operations, culminating in the preparation of large logical cluster states. Among the, in our opinion, most impressive and relevant achievements of this work, is the demonstration of a constant-entropy operation of a, in principle, scalable logical quantum processor which allows for the first time in any architecture the execution of deep universal quantum algorithms on tens of logical qubits. Whereas due to the limited size of the codes used the achieved logical qubit error rates do not yet reach the algorithmically interesting, few orders of magnitude lower logical error rates, this work provides all ingredients and lays out a very clear pathway to reach such regime. The work also provides and is in part enabled by significant improvements of decoding, in particular by the exploration of a series of increasingly powerful quantum erasure correction strategies. Overall, these results collectively represent substantial and truly impressive progress towards fault-tolerant quantum computing. The series of experiments, in particular the demonstration of deep universal logical qubit circuits constitutes clearly the most complex logical quantum algorithms demonstrated as of today.

The paper is overall written very clearly and accessibly. In particular the Methods sections provide not only valuable additional technical aspects on the experimental and theoretical (decoding, logical gate) techniques, but insightful additional discussion.

However, there are a few technical aspects that require clarification:

1. In the experimental demonstration of the lattice surgery based CNOT, only two rounds of measurements are carried out. This seems to be incompatible with fault-tolerance for the given code size, and the authors seem to try to compensate for this by post-selecting on the runs with equal joint logical operator measurement outcomes. What is the reason for not carrying out d rounds to ensure proper fault-tolerance of the protocol? Are there experimental limitations to the apparatus which currently prevent the authors from implementing the properly FT protocol? If possible, the authors should explore this experimental scenario. If for good reasons experimental data can't be taken to answer this question, what would the (numerically expected) performance look like for a properly fault-tolerant d-round protocol for given experimental error rates? The data shown in Fig. 3d, indicates a relatively high
surgery logical error of about 12%. How does this compare to distance-3 performance? Does the lattice-surgery CNOT operate below threshold, or above? If it is not below threshold, as it seems, the authors should comment on what the bottlenecks and required improvements would be to reach such regime. A clear and transparent discussion of this key point is needed, as the FT characteristics of a surgery protocol is important and more subtle than for transversal entangling operations. And for the community it would be highly valuable not only to learn that, not unexpectedly, lattice surgery performs worse than the transversal logical entangling operations. But in view of the time overhead required for atom shuttling, understanding whether the lattice surgery approach with static atomic arrays or lattices, in particular in the long run when logical cycle times will become more important than today, constitutes also a
viable approach for neutral atoms would be an important "architectural mechanism lesson" to learn.

2. There seems to be some ambiguity in the gate notation used throughout the manuscript. The experimental protocol likely implements CZ-gates, as for example illustrated in Fig. 3 or Fig. 4, but the plots show the number of CNOTs on the X-axis, and the text discusses the number of CNOT-gates per round. The authors write "CZ gates are realized as transversal CNOTs and Hadamards" in Figure caption 3, which suggests that CNOT-gates are used as native entangling gates. If this is the case, how are the Hadamard-gates on the surface code implemented?

3. The manuscript would benefit from a clearer presentation of the role and extent of postselection in the various protocols. It would greatly improve clarity if the authors could provide a summary, e.g. in form of a table, detailing the decoding strategy, the presence of postselection and its type (based on confidence, loss-information, non-matching stabilizer values), and how each of these factors contributes to the reported acceptance rate. A comment on why certain postselection types are employed, e.g. why is postselection based on the loss-information used for the color codes but not for the surface code, would help to follow and understand each protocol.

4. A comment on how fault tolerance is ensured for the various protocols would be a helpful addition. For example, how is the encoding circuit shown in Extended Data Fig. 10 made fault-tolerant (by e.g. a transversal copy-and-verification approach)? The section "Universality and synthesizing arbitrary unitaries" does not mention fault tolerance. The authors write that by leveraging transversal teleportation, it is "ensured that all physical errors are removed." How are all physical errors left behind, when (single) errors can propagate through transversal gates (onto single qubits in other blocks)? How is fault tolerance ensured, when e.g. each noisily encoded 3D color code block initially includes a single error on a data qubit? A more detailed discussion of these aspects would improve confidence in the robustness of the shown protocol.

5. In Extended Data Fig. 3, the authors explain that the non-destructive measurements take >10 ms, which is longer than a full round of QEC (4.45 ms, line 962). This seems to be a bottleneck for the execution of large scale quantum computations that rely on measurements. In the outlook and methods section (line 902 and following), the authors discuss possible experimental improvements for single- and two-qubit gates, but not measurements. It would be a valuable addition to the discussion if the authors could extend these parts and comment on how these slow measurements can be a limiting factor,

or why the authors believe that they are not.

6. The ML decoder used in this work requires extensive training on a given code size, followed by fine-tuning fed by experimental data. Working with a larger distance code (beyond d=5 explored in this work) seems to require re-training the respective neural networks entirely, with the training times and required samples probably not scaling favorably. Can the authors comment on the scalability of this approach?

7. A key achievement of this work is the implementation of error-corrected universal logical quantum gates and their exploration in deep logical circuits. Eventually, for logical algorithms to outperform their counterpart on physical qubits, it is important to compare all logical operations (in particular the gates, of course) with their physical counterparts. For the (transversal) Clifford operations, the logical operations seem to outperform the respective physical gates. How does the performance of, e.g., the up-to-depth 4, H-T teleported logical single qubit Bloch sphere dynamics shown in Figure 4 compare to the respective physical operations? Do they operate below break-even? If not, can the authors comments on or, better, numerically estimate which increased code distances or experimental parameter improvements are needed to reach such beneficial logical fidelities of the universal gate set over the physical operations?

In addition, there are some minor aspects, which should be addressed or clarified by the authors.

1. In Fig. 4, the authors prepare a logical state $|+\rangle_L$ and analyze the logical X-projection as a function of the rotation angle ϕ. However, the shown logical X-projection for $\phi = 0$ is -1, whereas one would expect $+1$ for an unrotated $|+\rangle_L$-state.

2. It would be helpful if the authors could add the fidelities or logical error rates (and how they are calculated) for the synthesized states presented in Fig. 4. Fig. 4c shows the angular deviation from the ideal target rotation, but this alone does not provide a direct measure of the logical fidelity or error rate associated with the prepared states.

3. In line 1219, the authors write “The curves are further normalized to highlight key features, with unnormalized data shown in ED Figure 9a.” What kind of normalization is done? 4. In the Methods Section (line 793), the authors write about the implemented lattice surgery protocol “A shot is counted as an error Z_L^1 Z_L^2 =–1 [...]”. For an input state $|{+}{+}\rangle$, the ZZ-measurement corresponds to a random projection and does not in itself imply an error. Is this additional postselection that is done and, if so, is it included in the stated acceptance rates?

The listed clarifications are required to improve the understandability of certain aspects. Overall, as outlined above, this work represents an impressive leap forward in the realization of fault-tolerant quantum computing. Therefore, we clearly recommend this manuscript for publication in Nature, provided the aspects outlined above are satisfactorily addressed.

References

[1] Rajeev Acharya, Dmitry A Abanin, Laleh Aghababaie-Beni, Igor Aleiner, Trond I Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Nikita Astrakhantsev, et al. Quantum error correction below the surface code threshold. Nature, 2024. doi:10.1038/s41586-024-08449-y.

[2] Nathan Lacroix, Alexandre Bourassa, Francisco JH Heras, Lei M Zhang, Johannes Bausch, Andrew W Senior, Thomas Edlich, Noah Shutty, Volodymyr Sivak, Andreas Bengtsson, et al. Scaling and logic in the color code on a superconducting quantum processor. arXiv preprint arXiv:2412.14256, 2024. doi:10.48550/arXiv.2412.14256.

[3] C Ryan-Anderson, NC Brown, CH Baldwin, JM Dreiling, C Foltz, JP Gaebler, TM Gatterman, N Hewitt, C Holliman, CV Horst, et al. High-fidelity teleportation of a logical qubit using transversal gates and lattice surgery. Science, 385(6715):1327–1331, 2024. doi:10.1126/science.adp6016.

[4] Ivan Pogorelov, Friederike Butt, Lukas Postler, Christian D Marciniak, Philipp Schindler, Markus Mueller, and Thomas Monz. Experimental fault-tolerant code switching. Nat. Phys., 21:298–303, 2025. doi:10.1038/s41567-024-02727-2.

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

Referee #5

(Remarks to the Author)

Referee Report for "Architectural Mechanisms of a universal fault-tolerant quantum computer" by Bluvstein et al.

Summary and Recommendation

The authors demonstrate an integrated neutral-atom architecture that combines a large number of primitives and architectural elements for fault-tolerant QEC: repeated surface-code QEC with loss-aware decoding, logical two-qubit operations via both transversal coupling and lattice surgery, universal single-qubit gate synthesis through teleportation, mid-circuit measurement, qubit re-use, and active entropy removal. Overall, the work is very impressive and largely delivers on its promise to "implement all key elements of a universal, fault-tolerant quantum processing architecture."

My overall recommendation is publication after some minor revision. The results are highly significant and relevant. My only suggested revisions are to add some clarification and, in a few places, some qualification about the claims.

Key strengths

1. The authors show clear "per round" below-threshold evidence with quantified gains from erasure signals. At d=5, they show a nearly 2x lower error rate per round than at d=3 using a four-round memory experiment. Crucial to these improvments was the improved decoding using machine learning and information about loss to augment the bare symdrome. I appreciated the error budget figure as well (2f in the Supplement), which shows that 3/4 of the total error is split between SPAM and CZ gates, with about half of that due to loss. The error budget breakdown certainly suggests that there are substantial gains to be had by addressing this information at the code level, and I'm very pleased to see the large performance gains the authors have made in this area. While this is not quite the most stringent standard for quantifying QEC logical performance (as the authors discuss), it is nonetheless impressive.
2. The authors consider a comparison between transversal gates and lattice surgery to see which is more practical. They quantify sensitivity to ancilla measurement errors and demonstrate that this may lead to inherent limitations with lattice surgery approaches (at least in some regimes). They then identify an optimum of ~3 transversal CNOTs per QEC round, matching a simple entropy-based model.
3. The authors demonstrate universal single-qubit logical gates via teleportation via the Reed-Muller code. This code has a transversal logical T gate, which the authors use to demonstrate universal logical gate synthesis, at least when the logical gates are transversal. They show how repeated teleportations generate increasingly dense images in the Bloch sphere when synthesized gates are interleaved with Hadamard rotations.
4. The authors tout the engineering advances that have improved mid-circuit measurement and re-use, for example the spin-to-position readout. It is not within my expertise to judge the difficulty of integrating these advances into the architecture, but these are clearly important advances for fault-tolerant operation.
5. The authors run deep circuits with constant internal entropy, which one can interpret as saying they are correcting errors at about the same rate as the errors are generated. Such a steady state is characteristic of operation in a fault-tolerant regime. They have also run repeated encoded cluster-state computations and shown constant stabilizer error and rapid decay of physical-error correlations.

Suggested revisions

1. I think it would be valuable to clarify the “threshold” language and long-round scaling. I found lines 120–129 lacking in sufficient detail in the main text, and potentially misleading. While the per-round result is strong, the text notes that the effective threshold can worsen by 1.15x with many rounds and 1.1x with 1 transversal gate per round. It would be valuable for the reader to include an explicit definition of “below threshold” used here, and (ii) scaling estimates or additional rounds (experiment or validated sim) showing whether the d-scaling persists when the number of rounds grows and when logic is interleaved. Finally, as the authors well know, there is a distinction between per-round improvements and per d rounds in a QEC cycle. I think it is a little late to first highlight this distinction in line 163, as then line 121 gives a misleading impression for what "below threshold" behavior means.
2. To help readers interpret Fig. 4, a brief note on the T-state generation/verification procedure with representative figures of merit (e.g., typical fidelity and success probability) would be useful. A concise sentence outlining the dominant error sources across the teleportation path and an order-of-magnitude indication of the space–time cost versus target rotation precision (even as a back-of-the-envelope estimate or pointer to Methods/SI) would nicely connect the demonstration to algorithmic overheads.
3. Since several claims use varying acceptance fractions (AF), a brief clarification of how AF influences results, particularly the reported “constant entropy” behavior, would strengthen the claim. I am a little nervous considering entropy in the context of postselection, since I think this could hide many potential distortions. One suggestion: A compact comparison at one fixed AF together with a sentence on reservoir-depletion limits and how continuous reloading would alter achievable depth would make the interpretation straightforward. But really any additional discussion on AF in the context of the constant entropy claims (in particular) are suggested.

Minor comments

1. Some phrases are a little awkward. Specifically "exponentially low, digitized errors" should maybe say low error *rate*. (Errors aren't "low".) Referring to generic 1Q gates as "analog-like" is a little misleading, especially given that they are implemented logically using digital gates. I appreciate what the authors are intending here, but I object to the term analog in this context, as I think it is misleading. The word analog in a similar context is repeated on line 203, and I think it is again being used incorrectly.
2. Figure 4a should mention explicitly in the caption that the traces are normalized. It is written in the axis, but it's not clear

what "Norm." means, and this should be stated in the caption.
3. Line 61 should perhaps say "polylogarithmic overhead" for precision.
4. You already list some suboptimal choices affecting fidelity (e.g., AOD intermodulation, SLM inhomogeneity, beam profiles, magnetic-field noise). A brief “expected near-term improvements” paragraph quantifying the projected logical-error gains from each item would be useful.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
I thank the authors for their thorough response to the questions and comments and all detailed revisions they made to the manuscript.

A few more comments regarding the responses to a-d:
(a) I agree that Steane-based QEC can work well provided a FT logical state preparation factory. Still, the word “robust” is slightly vague when discussing the fault-tolerance of a QEC protocol.
(b) The magic state distillation does start from an encoding circuit, but is not guaranteed to prepare an error-free target logical state (that is why magic state distillation is required). And for magic state distillation, FT error correction on magic states is required before creating the concatenated code from non-perfect magic states.
(c) I agree that one can achieve arbitrary angle rotation by switching between [[15, 1, 3]] and [[7, 1, 3]] codes. This is based on the (one-directional) transversal CNOT gate between the two codes though, which is specific to the choice of codes.
(d) I agree that transversal CNOT can perform logical teleportation, but then single-qubit errors would propagate through layers. The alternate approach requires transversal H, which is, again, specific to the choice of codes.

In general, I would still think this work is closer to “realizing several key ingredients of FTQEC using different, suitable codes” rather than “realizing FTQEC”, so I thank the authors for softening their claim on the key aspects of this work. The code switching protocol between [[7, 1, 3]] and [[15, 1, 3]] can, in principle, realize FTQEC, based on these specific transversal gate gadgets: transversal CZ, one-directional transversal CNOT, transversal T in [[15, 1, 3]] and transversal H in [[7, 1, 3]]. This transversal gate set is so specific such that it is hard to imagine that for a general code switching protocol all those gates would work, and we have good reasons to believe that code switching between [[7, 1, 3]] and [[15, 1, 3]] is not the ultimate solution to practical FTQEC due to the low error threshold. Nevertheless, I can imagine “realizing FTQEC by switching between [[15, 1, 3]] and [[7, 1, 3]]” to be the next step of the authors’ work, and this work indeed provides many, although not all, tools towards that goal. I therefore recommend this work to be published in Nature.

Referee #2

(Remarks to the Author)
The authors have carefully addressed my comments. Together with addressing the comments by the other reviewers, this has significantly improved clarity about key aspects of what has and what has not yet been achieved in this work. I am therefore happy to recommend publication of this very impressive advancement towards scalable fault-tolerant quantum computing.

I still have two comments I ask the authors to consider:

I thank the authors for the inclusion of the newly added Extended Data Fig. 8d, on the numerically simulated expected performance of the lattice surgery protocol for d=3 and d=5 for different numbers of stabilizer measurement rounds. This increases clarity regarding the choice of the number of rounds and of the relative performance for distance 3 and 5. Still, I think for clarity, it is important to highlight more explicitly that the surgery protocol with only 3 rounds (instead of 5) is not fault-tolerant in the strict sense as required for a fault-tolerant quantum circuit for such entangling operation via surgery on the d=5 code. I suggest to slightly rephrase the corresponding last sentence in the figure caption accordingly, to read e.g. “Whereas the protocol with 3 rounds for d=5 is not strictly fault-tolerant, we observe that the minimum error occurs at roughly 3 rounds for d=5, as opposed to 5.”

In the newly added subsection on “Small-angle synthesis”, the sentence in line 1117 “By adding the appropriate feedforward at each teleportation step, this protocol becomes deterministic.“ is slightly ambiguous and could be misinterpreted as if the authors did this feedforward (experimentally). I understand that this is not what the authors want to express, but for clarity, to avoid such potential misconception, I suggest rephrasing the sentence slightly as “Adding the appropriate feedforward at each teleportation step would render the protocol deterministic.”

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

Referee #5

(Remarks to the Author)
I have reviewed the authors' reply and their changes, and I am satisfied that the current article meets Nature's standards.

Referee responses and summary of revisions

Nature Manuscript number: 2025-06-15475 Bluvstein

September 6, 2025

We would like to thank all Referees for their careful reading of our manuscript and many useful comments and suggestions. In what follows we address all comments point by point and indicate revisions when appropriate. The blue text are the Referee comments, the black text are responses, the red text are revisions, underlined text indicates text additions, and strike-throughs indicates text removals. At the editor's request we have made efforts to not lengthen our manuscript, and have shortened the abstract to the requested length.

Reviewer: 1

The manuscript by Bluvstein and Geim et. al. demonstrates several building blocks towards a practical logical quantum processor with fault-tolerant universal operations. They first implement projection onto logical qubit subspace via stabilizer measurement, successfully demonstrate logical error suppression and compare different logical entangling methods using surface codes. They then implement synthetic arbitrary-angle single-qubit rotations with the [[15, 1, 3]] code, and finally explore logical-level deep-circuit behavior using logical teleportation with [[7, 1, 3]] and [[16, 6, 4]] codes. Several hardware updates, including spin-to-position non-destructive readout and loading from a reservoir, are employed for better hardware performance and faster clock speed. The results are solid, revealing a promising roadmap towards universal fault-tolerant quantum computation using reconfigurable neutral atom arrays.

We thank the Referee for their positive evaluation of our manuscript.

The authors declare that in this work they "implement all key elements of a universal, fault-tolerant quantum processing architecture". Considering all the impressive results in this work, I would still find this claim slightly exaggerative, with reasons listed as follows:

We thank the Referee for these insightful questions. While we agree with the Referee that there are several viable approaches to FTQC, in our work we have

chosen the protocols that are perhaps the simplest and clearest to enable universal FTQC. We first address a-d point-by-point, and then answer the question of how they all come together.

a. General fault-tolerant stabilizer measurement is missing — the NZNZ pattern for surface code is FT, but would not work for codes such as [[7, 1, 3]] and [[15, 1, 3]] which are used in logical non-Cliffords and logical teleportation below. Shor/Steane/Flag type stabilizer measurement protocol is in general required.

We agree with the Referee that practical fault tolerance properties depend on both specific circuit and codes involved, and need to be carefully examined for each specific system. Although non-flagged measurements may be not-strictly-FT for d=3 codes, for topological families, i.e., 2D and 3D color codes, they can still be practically sufficient especially for codes of larger distance. This is amplified by the fact that hook-type errors can only spread to a weight-2 error and correspond to X-type errors which are more rare in neutral atom systems, as is briefly discussed in the Methods section *"NZNZ stabilizer gate pattern"*.

We further note that various methods of syndrome measurement based on transversal gates are also valid for FT syndrome extraction as demonstrated in our and other previous works using Steane-based QEC on color codes.

Revision: To clarify these points, we have added the following to the Methods discussion *"NZNZ stabilizer gate pattern"*

"In our experiments in particular, we choose an ordering of $N\ Z\ Z_r\ N_r$, where r represents performing the reverse ordering (see ED Fig. 6d). In addition to such structuring helping preserve fault-tolerance against hook errors, we also note that the dominance of Z-type errors means that most errors do not lead to propagated errors between the middle two CZ gate layers. Due to these reasons, in simulations we do not observe having spatially alternating "N" and "Z" patterns helps performance (not plotted). Although the $d = 3$ color codes studied here could also suffer from hook errors under repeated stabilizer measurement, we similarly expect that they would be robust to these errors with increased code distance. Moreover, we note that as studied in Ref. 11, Steane-style QEC can be effectively used for fault-tolerant syndrome extraction on color codes in neutral atom systems."

b. Mid-circuit error correction after decoding is missing — The authors claim that +1-logical subspace is required for non-Clifford gates, but it is not demonstrated.

In our experiments, a unitary encoding circuit is used to ensure stabilizers are deterministically prepared in the +1 logical subspace for non-Clifford gates. This is similar to what happens in a magic state distillation encoding.

Revision: To emphasize this point, we have added

"Figure 4a shows experimental measurements where we use encoding ~~quantum~~ circuits (see Methods) to deterministically prepare ~~create~~ 2D color codes (Steane codes), and 3D color codes (Reed-Muller codes), and subject them to global phase rotations φ around the Z-axis."

We do, however, agree that mid-circuit correction is important in key contexts. In particular, fixing atom loss is a key aspect that cannot be corrected in software. In Figure 5 and 6, we demonstrate the realization of deep circuit operations with mid-circuit / feedforward atom loss correction in order to maintain constant entropy behavior.

c. General non-Clifford operations are missing — The synthetic arbitrary-angle rotation using [[15, 1, 3]] code is based on the specific transversal universal gate set of the code (more specifically, transversal CZ available), and is not guaranteed to work with other codes.

We agree that the [[15,1,3]] protocol used here is more compact due to the use of transversal T and transversal CZ. However, we emphasize that this protocol uses the same underlying mechanisms that one would use for magic state distillation.

We further note that, while the CZ gate is convenient, it is not required in our approach. For example, in ED Fig. 9 we study code-switching-based implementations of non-Cliffords on a 2D code, where the 15-qubit code provides T-state creation which is then teleported onto the 2D code via a transversal CNOT. In ED Fig. 9e, we decompose this CNOT into a CZ gate and Hadamards to illustrate the equivalence with the protocol for small-angle synthesis using the [[15,1,3]] and therefore highlight that the transversal CZ is not required. Broadly, doing a CNOT onto a 2D code permits Hadamard gates which also permits the universality (see also Response to Referee 2, Comment 2).

Revision:

"ED Fig. 9 explores universality with transversal CNOTs and code-switching ~~code-switching and transversal gates~~ between 2D and 3D codes, again finding the underlying source of universality is the

measurement of the 3D code."

d. The logical teleportation protocol for cleaning up physical-level errors is also based on transversal CZ gates.

Similar to the point above, the CZ gates are not required and can instead be realized with CNOT gates. I.e., here we prepare all codes in $|+_L\rangle$ and do CZ gates, but equivalently we could alternate codes in $|+_L\rangle$ and $|0_L\rangle$ and then realize CNOT gates, or use transversal Hadamard gates. This recovers MBQC-type repeated QEC.

Revision: To highlight this equivalence we have made the following addition to the Methods section *"Use of teleportation for ensuring error removal."*

"We note that the same behavior can be achieved in codes with transversal CNOTs with appropriate preparation of basis states or transversal Hadamards, like in conventional MBQC with surface codes, and is not predicated on having a transversal CZ gate in the code."

All those are parts of my major concern regarding the results: How do all these building blocks come together to perform a FT logical quantum processor using codes with decent encoding rate/distance, but without built-in good transversal gate set (qLDPC for example, which may only have transversal logical Paulis and CNOTs)? While the claim is overall not hurting the solidity of the results, I would like to see the authors' thoughts on the steps forward before I can fully recommend this work to be published in Nature.

As our responses above indicate, we believe the approaches chosen here highlight a simple and clear implementation of universal FTQC mechanisms. We further emphasize that these core mechanisms are also applicable to high-rate-code settings. For instance, we demonstrate that the same principles apply to physical-level-error clean-up for high-rate [[16,6,4]] codes in Figure 6. While the teleportation-based small-angle synthesis could also be used here, we agree with the Referee that finding efficient methods of realizing non-Clifford operations on high-rate/qLDPC codes is an outstanding challenge for both theory and experimental implementations.

Revision: To this end, we have added to the end of the Methods section *"Universality, transversal gates, and Eastin-Knill."*

"While these techniques are easily understood in the context of topological codes, they can be also utilized for qLDPC or general high-rate codes. In particular, as demonstrated in Figure 6f, the teleportation protocols are effective for high-rate codes, and in principle,

teleportation-based small-angle synthesis could be also used here as well. At the same time, efficient, parallel generation of logical magic in such high-rate codes is an outstanding problem and an active area of theoretical research."

We have further modified the abstract to read

"Here we utilize reconfigurable arrays of up to 448 neutral atoms to implement ~~all~~the key elements of a universal, fault-tolerant quantum processing architecture and experimentally explore their underlying working mechanisms."

Besides that, I have several more questions and suggestions for the authors to consider before resubmitting their work:
1. For below-threshold performance:
a. What is the source of that 8% detector probability with post-selection (Fig. 2b and 2c)? An error budget would help with better understanding of the hardware performance.

Figure 2f presents a simulated error budget that agrees well with our experimental results and separately measured experimental error sources, which indicates our approximate experimental error contributions. The plotted contributions are incorporating loss, but in the absence of any loss errors (i.e., ignoring SPAM loss and CZ loss) the simulated error budget in Fig. 2f predicts $\approx 6.2\%$ detector error arising from single qubit errors, T_1/T_2, SPAM, and CZ Pauli error. While our overall measurements are consistent with the simulated error budget (Extended Data Figure 6), the loss-postelected case slightly underestimates the observed loss-postselected detector probability of 8.06(4)%. We nevertheless expect the distribution of errors is still approximately correct.

Revision: We have modified the caption of Extended Data Fig. 6f to read

"Comparison to simulation. Left: detector error probability, converting loss to qubit state 0. Right: detection correlation matrix p_{ij}. Both metrics show good agreement between simulation and experiment in the structure and magnitude of the errors. Simulation reproduces experimental detector error probability, being 10.71(3)% (experiment) and 10.5% (simulation) for the "detect loss" metric. With loss postselection the simulation slightly underestimates the loss-postselected detector error (8.06(4)% compared to ≈6.2%).

b. I find the explanation of the four metrics in Fig. 2(b)(c) a bit ambiguous:
i. Are the atom loss information from a final readout of the data block, or from

observing the flickering pattern of stabilizers?
ii. Are the four groups of detector loss data extracted from the same experimental dataset and the only difference is the data processing?
iii. If the answer to the previous problem is yes, then I don't understand the difference between "bare" and "detect loss": doing an entangling gate with an atom in $|0\rangle$ state is conceptually different from doing an entangling gate with nothing (atom loss). It is confusing how to count loss as state $|0\rangle$ or not in the data processing under a pre-determined physical process.
iv. If "post." is post-selecting on no atom loss, then where does the "post." data with non-zero loss rate come from?

We thank the Referee for pointing out these potential ambiguities.

(i) The atom loss information is directly obtained from the final readout of the data block.

Revision: we have modified the caption of Fig. 2b to read

"Detector error probability as a function of measured data qubit loss in each shot, analyzed by partitioning the total dataset."

(ii) The same dataset is used throughout Figs. 2b and c and the difference is indeed only in the data processing.

(iii) We agree that treating a lost atom as state $|0\rangle$ in data processing is not the same as the atom being physically in $|0\rangle$ during the computation. Here, "bare" calculates the detector error probability as if the qubit readout did not include loss detection, in which case lost atoms are necessarily misidentified during measurement as one of the qubit states (arbitrarily chosen to be $|0\rangle$). In contrast, "detect loss" uses the loss detection to exclude detectors whose value is not known due to a lost atom.

Revision: we have added the following to the Methods section "*Additional experiment and data analysis details*"

"The four metrics are (i) 'bare', where loss is converted to qubit state 0, effectively corresponding to no loss detection (ii) 'detect loss', where projective measurements whose value is 'loss' are not counted erroneously (iii) 'supercheck', where stabilizers with a lost data qubit are formed into superchecks for all prior rounds, and (iv) 'postselected', where detectors involving any lost atoms are ignored."

(iv) "Post" only postselects on no atom loss within individual detectors, and not across the entire code. For non-zero loss rate, the postselected detector error is therefore the average across all detectors in the shot which do no touch a lost

data or ancilla qubit.

Revision: To emphasize this, we have made the following change to the caption

"'Bare' counts loss as state $|0\rangle$, 'detect loss' does not make this erroneous assignment, 'supercheck' multiplies detectors around lost atoms, and 'post.' postselects on all atoms of ~~a detector~~ individual detectors being present."

c. For ML decoder, do the authors have an estimation of how the time complexity scales with the code size?

We thank the Referee for this question. With certain models, for example the artificial neural network studied in Ref. 96, the inference time can be made approximately independent of code distance, d. On the other hand, in attention-based architectures such as Ref. 32, the decoding time increases approximately quadratically with d. The scalability of different ML decoders along many axes - including speed, accuracy, memory requirements, and training complexity - is an important open research question, and some of these issues will be addressed in our forthcoming publication (Ref. 51).

Revision: At the end of the method section *"ML decoder - surface code"* we have added

"Although the ML decoder used here is not directly scalable to high distance codes – for example requiring re-training for each code distance and specific circuit, with the number of training samples growing exponentially – exploring different extensions of such neural network architectures for scalable decoding is an exciting direction of ongoing research [Ref. 32, Ref. 51, Ref. 96]."

2. General comment: For several sets of data analysis, the choices of acceptance fractions are vaguely clarified. Detailed explanations would be helpful.

We thank the Referee for this suggestion and have added a summary of the types of decoder and acceptance fractions employed.

Revision: We have added a file to the Supplementary Information tabulating all decoding and methods of error correction and error detection, titled

Summary of decoders

To further clarify the choices of acceptance fractions, we have added the following to the Methods section *"Additional experiment and data analysis details"*

"We use a lookup table for decoding and plot all three 3D color code curves with an acceptance fraction of 46% and the 2D color code with 74%, corresponding to a rescaling by the number of physical qubits in the code. This slightly sharpens the plateaus, which are otherwise smoothened in the presence of logical Z errors."

and added a reference to the Methods to clarify the use of acceptance fractions in Figure 6:

"On the logical-qubit level we observe the expected correlations between logical outcomes corresponding to successful algorithm evolution, with a decay with distance corresponding to an effective algorithmic error rate that improves with decreasing entropy (corresponding to stricter postselection, Methods ~~lower acceptance fraction~~)."

3. For the non-destructive readout, it seems appropriate to cite other approaches in the community that have been used to demonstrate QND readout quite a while ago in some cases. For instance, stretched readout techniques with alkalis were demonstrated by Meschede and Saffman in 2017 and are used in large-scale alkali processors by Saffman and Pritchard, QND readout of optical clock qubits was realized and exploited by Endres and Kaufman in 2019, and QND readout of nuclear spin qubits was realized and exploited by Covey and Atom Computing in 2023. As such, the novelty of this capability here seems to be somewhat overstated, as only the Weiss approach that was essentially duplicated here is mentioned. That said, I appreciate the demonstrations of the significance of QND readout for deep circuits.

We thank the Referee for the suggested references and agree that many complementary approaches to QND readout have already been explored by the community. For ^{87}Rb in particular, stretched readout based on cycling transitions requires deep traps to retain the atom (e.g., 3.5 mK by Meschede in 2017, compared to typical $\sim$0.5 mK in these experiments) and is therefore challenging when limited by available laser power. Alternatively, simultaneous cooling can reduce trap depth requirements at the cost of longer readout times (typically tens of ms). Note that we do not intend to overstate the novelty of our spin-to-position conversion which is inspired by the Weiss approach although differs in practical details; for example, we combine the spin-selective pinning potential with tweezer confinement rather than solely relying on lattice trapping.

Revision: In the Methods section *"Spin-to-position conversion with a one-dimensional optical lattice"*, at the end we have added

"We remark that non-destructive qubit readout has previously been realized using stretched states [Meschede2017, Saffman2017], but however also requires several times deeper traps than the present approach. More recently, related protocols for fast, high-fidelity readout have been realized across several platforms [Norcia2019, Huie2023, arXiv 2507.01720, arXiv 2507.01011]."

4. As a related point, the community would benefit from must faster QND readout. The readout time does not seem to be directly mentioned, but it appears to be 3-10 ms depending on the settings. The authors should comment on the prospects to bring this down to the ∼10-100 μs level. This also goes for cooling and reset. Given that destructive measurement can be performed in ∼10 μs, it may be appropriate for the authors to briefly present their vision for whether slow non-destructive measurement is better than fast destructive measurement, particularly with the advent of coherent reloading to obviate atom loss constraints.

We agree that faster QND readout, cooling, and reset are highly desirable. Although slower than destructive measurements, we find that non-destructive readout is crucial for improved QEC performance by providing loss information. Even with coherent reloading - required for continuous operation over long computations - erasure information and efficient re-use of resources will likely remain central to neutral atom architectures. Moreover, readout and computation can often be parallelized, for example by delaying Pauli feedforwards until the next non-Clifford operation. We estimate that the cycle rate demonstrated here can be increased by nearly an order of magnitude through simple optimizations, as discussed in the Methods section *Processor clock speed*. Additional improvements are possible, for example faster reset ($\sim 20\mu s$, Ref. 53), optimized atomic motion [Ref. 58], and rapid imaging as demonstrated in other species (see above).

Revision: We have added a discussion of readout speed to the Methods section "*Processor clock speed*":

"In the repeated Rabi calibration in Fig. 5b, we did optimize for speed and achieved a cycle time of 4 ms. Although the imaging here was global as a demonstration of fast calibration, we expect such speeds can also be achieved in a zoned manner. Destructive measurement can be faster than the qubit re-use approach employed here, but the absence of loss information degrades QEC performance (Fig. 2) and further increases required qubit reloading rates for continuous operation. While non-destructive readout is slower, this may or may not bottleneck operations depending on, e.g., when the next non-Clifford operation occurs. Comparing these holistically in the context of

a whole architecture is an important avenue of future research.

5. In Figures 1c & d, presumably the "QEC" refers only to stabilizer measurement and not active correction via conditional operations. Although the passive correction or conversion due to the Zeno effect is very nice, the language surrounding this study is somewhat misleading. If the authors were to actually do syndrome extraction and correction in each "QEC" round, to what extent is it meaningful to "track the state of the system" to detect incoherent errors? Perhaps coherent errors would be corrected on each round, and then incoherent errors would be inferred based on comparing several rounds?

We thank the Referee for this question. Here, QEC refers to both stabilizer measurement and the subsequent in-software tracking of detected Pauli errors, which together constitute syndrome extraction and correction. Since there are no non-Clifford operations, active correction of physical X and Z errors is not required (see Methods, *"Feedforward in universal processing."*). We distinguish this in-software correction ('tracking') of bit- and phase-flip errors from the conversion of coherent errors into incoherent ones by the projective measurement ('Zeno effect'). As the Referee correctly states, coherent errors are projected in each round and these now-incoherent errors (alongside native incoherent errors) are identified by comparing stabilizers between adjacent rounds to form detectors.

Revision: We have updated the caption of Fig. 1c,d

"Repeated ~~correction~~ stabilizer measurement reduces error build-up through both the Zeno effect and error tracking between rounds (right plot is at a fixed $\theta/2\pi = 0.016$)."

6. For transversal CNOT vs. lattice surgery, it is confusing that the transversal CNOT error rate does not change at all at various ancillary measurement errors (Fig. 3c). Is there no error correction (even in software) required after stabilizer measurements if the measurement is just for projecting to logical subspace/remove entropy? Is it some properties related to correlated decoder?
7. As a related point, I find it less compelling that the "data" is entirely from post-processing. What role is the experimental data actually playing here? Is the experimental data used shot-by-shot, or is it averaging over many realizations with the post-processing (injecting measurement errors) on averaged data? Why is this not just a theory result in terms of what error probability per logical operation we expect for a given ancilla measurement error? Or at least, why are there not theory lines on Fig. 3(c)? It seems like the role of real data is to set the offset (error probability per logical operations when there is no injected ancilla measurement errors), but the trend should presumably entirely match a simple theory.

There is a small change in the transversal CNOT case, where the error rate rises weakly from 1.5% to 2.0%. Here, error correction is indeed done in-software. However, as the Referee correctly points out, because the stabilizer measurements serve to remove entropy and are not used for logic, the measurements are significantly more robust assuming the appropriate correlated decoding is executed. It is indeed related to the correlated decoder as the correlated decoder ensures the appropriate logical observables are decoded in a fashion that is robust to time-like errors [Cain et al 2025], so that measurement errors only degrade the performance of the entropy removal but do not corrupt the logic gate itself.

The experimental data is used shot-by-shot and not as averaged quantities: in each shot, we first flip the measurement of each ancilla qubit with probability p (where flipping a lost atom does nothing). The modified measurements are then decoded with the MLE decoder, increasing the ancilla measurement error in the noise model appropriately.

Revision: We have added further details of the error injection protocol to the Methods section "*Additional experiment and data analysis details*"

"In Fig. 3c, the transversal CNOT is shown for 3 CNOTs per QEC round~~, and the injected~~. For each experimental shot, measurement error is applied in-software with probability p independently to every ancilla qubit (where an error on a lost atom does nothing). The modified measurements are then decoded, with the ancilla measurement error in the MLE noise model increased appropriately."

8. As a related point, how would this story change if d rounds of measurements were used for lattice surgery instead of just one? It seems straightforward to implement this in post-processing and straightforward to have a theoretical expectation. To be clear, I am not suggesting that lattice surgery is better than transversal gates, but the presented studies seem somewhat contrived.

We thank the Referee for this question. In these experiments, we first prepare the two d=5 codes in $|+\rangle_L$ to reliably initialize the X stabilizers and X logical operators in their +1 eigenstates. For the transversal state preparation, the Z stabilizers do not have to be deterministically prepared (as studied in theoretical works on correlated decoding, e.g., Ref. 12). To perform lattice surgery, we use two rounds of ancilla measurement to measure $Z_L^1 Z_L^2$ given by the middle three ancillas which merge the two codes. As the Referee states, one typically assumes d rounds is required for fault-tolerance since the ancilla measurements must be correct. Here, as discussed in the Methods section *"Additional experiment and data analysis details"*, we find in simulation that approximately 3 rounds is optimal for our experimental error model, due to the error associated

with syndrome extraction. By using error-detection to additionally compensate for the reduced number of rounds, we approximately recover optimal performance.

Revision: To accompany the discussion in the Methods, we have added a new subfigure, **ED Fig. 8d**, showing the results of lattice surgery simulations for up to d rounds of measurement using our experimental error model. We have updated the caption to read

"d, Numerical simulations of lattice surgery on two rotated surface codes initialized in $|+_L\rangle$ undergoing up to five rounds of repeated stabilizer measurement with gate ordering taken from N, Z, Z_r, N_r, N. We simulate the same protocol studied in Figure 3 to measure the logical ZZ product, using an experimental error model with loss, and plot the logical error probability in the obtained Bell state. At two rounds, simulation (13.1%) approximately reproduces experiment (15.2(3)%). The minimum error occurs at roughly 3 rounds for d=5, as opposed to 5."

and added a reference to this figure in the relevant Methods section

"We find in numerical simulations using our experimental error model that the optimal number of QEC rounds for this circuit is approximately 3 (as opposed to 5), and that by using error detection with 2 rounds we recover a similar performance to this optimal value found in numerics (ED Fig. 8d)."

9. The issue of feedforward comes up again for the synthesis of T states. Since the feedforward operations are only being applied in software, true T states are presumably only generated probabilistically via post-processing. As such, the authors should comment on the "yield" and how it scales with code distance. They should also comment on whether such teleportation protocols are deterministic (when logical feedforward is applied) and how scalable this approach is to larger codes. It may also be appropriate to provide a fidelity for T state preparation. The note about ED fig. 9f and the CHSH inequality suggests that the fidelity is several 9's, but the discussion in the main text is surprisingly qualitative.

As discussed above, in our approach feedforward is not required for synthesis of the T states - this is realized simply by preparing a $|+_L\rangle$ Reed-Muller code deterministically and then realizing a deterministic transversal T gate. With our deterministic T states, we then teleport them to realize HTHTH-type small-angle synthesis. While feedforward is indeed required to achieve deterministic rotations, since in this work we attempt to make all angles to show the exponential coverage on the Bloch sphere, we did not apply these feedforward operations

in-hardware, and the specific angle generated is dependent on the outcome of the logical measurement.

Nevertheless, while with feedforward the protocol is deterministic, as mentioned in the caption the plotted data uses error detection to improve the performance and highlight the Solovay-Kitaev physics and small-angle synthesis. As an example, in ED Fig. 9 we perform an error-corrected Bell test by making a Bell state between a deterministically-prepared T state on the Reed-Muller code and a Steane code. The measured CHSH inequality saturates the quantum bound, with error detection used to improve the fidelity (as indicated in Figure caption).

Revision: To address the above comments, we have added a subsection under *"Methods of universality"*:

"Small-angle synthesis"

Additionally, to emphasize that the generation of $|+_L\rangle$, and hence T states, is deterministic, we have updated the text to read

"Figure 4a shows experimental measurements where we use encoding ~~quantum~~ circuits (see Methods) to deterministically prepare ~~create~~ 2D color codes (Steane codes), and 3D color codes (Reed-Muller codes), and subject them to global phase rotations φ around the Z-axis [Ref.13]."

Reviewer: 2

Report on manuscript Architectural mechanisms of a universal fault-tolerant quantum computer

In their work, the authors present an experimental demonstration of all essential components required to perform universal error-corrected fault-tolerant logical quantum computation on multiple logical qubits. This is enabled by a number of key improvements to the experimental architecture, as compared to previous work by earlier work by the team demonstrating a logical quantum processor. Specifically, the authors introduce a technical upgrade of their neutral atom quantum processor, transitioning from spin-to-loss to spin-to-position conversion. This enables non-destructive state readout and additional loss detection at the end of a computation. On this platform, the authors demonstrate repeated rounds of QEC on the rotated surface code of distance d= 3 and d= 5, operating with approximately a factor of 2 below threshold performance. They extract logical error rates for up to five rounds of quantum error correction (QEC) and achieve a similar error-suppression per round as shown in [1, 2]. Experiments are accompanied by numerical simulations, which allow the authors to estimate the relative error contributions of different physical error mechanisms. Building on their QEC protocols, they implement logical entangling gates by means of transversal CZ-gates interleaved with QEC. They find an optimal logical error per logical operation for three CZ-gates per QEC round, and conclude that the entropy introduced by the logical entangling gates is exactly balanced by the stabilizer measurements. They further realize logical entangling operations through lattice surgery, for the first time demonstrated in a neutral-atom processor, by means of joint logical ZZ-measurements on an initial state $|++\rangle$ and achieve error rates comparable to recent works [3]. The authors then prepare distance-3 2D and 3D color codes in $|+\rangle_L$ and observe robust angles under rotation about the Z-axis, which correspond to stabilizer revivals. Arbitrary single-qubit rotations are then synthesized on the 3D color code with transversally realized teleportation-based protocols, where the Eastin-Knill theorem is circumvented with (non-unitary) logical measurements. Based on these protocols, the authors inject layers of logical HT-gates, thus preparing many states on the Bloch sphere that cannot be reached with purely unitary transversal operations by any single QEC code, conceptually similar to earlier demonstrations with trapped ions [4]. Furthermore, the authors implement deep logical circuits comprising of up to 27 layers of transversal entangling operations on 32 instances of the distance-3 2D color code. They leverage teleportation protocols such that physical errors are removed while the logical information propagates through the circuit. For the prepared logical 1D and 2D cluster states, they can preserve correlations substantially longer than for physical qubits. Lastly, the authors encode high-rate codes and, again, execute transversal entangling gates to prepare large cluster states and relialy teleport logical quantum states forward in a measurement-based QC fashion.

The key milestones include repeated rounds of quantum error correction (QEC), logical entangling gates, a protocol for arbitrary single-qubit logical operations, culminating in the preparation of large logical cluster states. Among the, in our

opinion, most impressive and relevant achievements of this work, is the demonstration of a constant-entropy operation of a, in principle, scalable logical quantum processor which allows for the first time in any architecture the execution of deep universal quantum algorithms on tens of logical qubits. Whereas due to the limited size of the codes used the achieved logical qubit error rates do not yet reach the algorithmically interesting, few orders of magnitude lower logical error rates, this work provides all ingredients and lays out a very clear pathway to reach such regime. The work also provides and is in part enabled by significant improvements of decoding, in particular by the exploration of a series of increasingly powerful quantum erasure correction strategies. Overall, these results collectively represent substantial and truly impressive progress towards fault-tolerant quantum computing. The series of experiments, in particular the demonstration of deep universal logical qubit circuits constitutes clearly the most complex logical quantum algorithms demonstrated as of today.
The paper is overall written very clearly and accessibly. In particular the Methods sections provide not only valuable additional technical aspects on the experimental and theoretical (decoding, logical gate) techniques, but insightful additional discussion.

We thank the Referee for their positive evaluation of our manuscript.

However, there are a few technical aspects that require clarification:
1. In the experimental demonstration of the lattice surgery based CNOT, only two rounds of measurements are carried out. This seems to be incompatible with fault-tolerance for the given code size, and the authors seem to try to compensate for this by post-selecting on the runs with equal joint logical operator measurement outcomes. What is the reason for not carrying out d rounds to ensure proper fault-tolerance of the protocol? Are there experimental limitations to the apparatus which currently prevent the authors from implementing the properly FT protocol? If possible, the authors should explore this experimental scenario. If for good reasons experimental data can't be taken to answer this question, what would the (numerically expected) performance look like for a properly fault-tolerant d-round protocol for given experimental error rates? The data shown in Fig. 3d, indicates a relatively high surgery logical error of about 12%. How does this compare to distance-3 performance? Does the lattice-surgery CNOT operate below threshold, or above? If it is not below threshold, as it seems, the authors should comment on what the bottlenecks and required improvements would be to reach such regime. A clear and transparent discussion of this key point is needed, as the FT characteristics of a surgery protocol is important and more subtle than for transversal entangling operations. And for the community it would be highly valuable not only to learn that, not unexpectedly, lattice surgery performs worse than the transversal logical entangling operations. But in view of the time overhead required for atom shuttling, understanding whether the lattice surgery approach with static

atomic arrays or lattices, in particular in the long run when logical cycle times will become more important than today, constitutes also a viable approach for neutral atoms would be an important "architectural mechanism lesson" to learn.

We thank the Referee for these insightful questions. In the surface code experiments, a degrading trap laser limited the number of qubits we could use and therefore we only implemented two rounds of stabilizer measurement. However, in simulation we find that approximately 3 rounds is optimal for our experimental error model, as opposed to the conventional 5 rounds, due to the additional error associated with syndrome extraction (noted in Methods, *"Additional experiment and data analysis details"*). The simulation predicts a logical error of 11.6% after three rounds (and 13.1% after two), comparable to the 11.8% obtained from experiment with error detection. We further simulate the same protocol for d=3 and find that the lattice surgery here operates slightly above threshold.

Revision: To accompany the discussion in the Methods, we have added a new subfigure, **ED Fig. 8d**, showing the results of lattice surgery simulations on d=3,5 surface codes for up to d rounds of measurement using an experimental error model. We have updated the caption to read

"d, Numerical simulations of lattice surgery on two rotated surface codes initialized in $|+_L\rangle$ undergoing up to five rounds of repeated stabilizer measurement with gate ordering taken from N, Z, Z_r, N_r, N. We simulate the same protocol studied in Figure 3 to measure the logical ZZ product, using an experimental error model with loss, and plot the logical error probability in the obtained Bell state. At two rounds, simulation (13.1%) approximately reproduces experiment (15.2(3)%). The minimum error occurs at roughly 3 rounds for d=5, as opposed to 5."

As the Referee notes, fault tolerance is more subtle for lattice surgery than transversal operations. In particular, in transversal operations (with correlated decoding) only space-like errors can occur and the decoding of the logical operators behaves as a memory experiment, as studied in Ref. 12. In lattice surgery, time-like errors can additionally occur since the ancilla measurement values must be correct to fault-tolerantly perform the logic operation. This is what leads to the higher logical error and the sensitivity to ancilla measurement error studied in Fig. 3. Finally, as discussed at the end of the Methods section *"Processor clock speed"*, the $O(d)$ reduction in number of QEC rounds with transversal operations more than offsets the time overhead for atom shuttling. E.g. for large-scale, practical algorithms, it results in a factor of 10-20 slower logical cycle time than in static, planar architectures with much faster physical clock speed of $1\mu s$ (as appropriate, e.g., for superconducting processors).

Revision: In the Methods section *"Additional experiment and data analysis details"*, we have added a discussion to text

"We find in numerical simulations using our experimental error model that the optimal number of QEC rounds for this circuit is approximately 3 (as opposed to 5), and that by using error detection with 2 rounds we recover a similar performance to this optimal value found in numerics (ED Fig. 8d). While transversal operations are only sensitive to space-like errors with correlated decoding, additional time-like errors in lattice surgery lead to the higher logical error and the sensitivity to ancilla measurement studied in Fig. 3c."

We have also added a new Reference in the Methods section *"Processor clock speed"*:

"As such, we estimate the present methods are slower by a factor of $\sim 10-20$ relative to a conventional planar architecture, associated e.g. with superconducting qubits, with 1 μs speed per stabilizer measurement cycle [Gidney2021, Acharya2024, Zhou2025]."

2. There seems to be some ambiguity in the gate notation used throughout the manuscript. The experimental protocol likely implements CZ-gates, as for example illustrated in Fig. 3 or Fig. 4, but the plots show the number of CNOTs on the X-axis, and the text discusses the number of CNOT-gates per round. The authors write "CZ gates are realized as transversal CNOTs and Hadamards" in Figure caption 3, which suggests that CNOT-gates are used as native entangling gates. If this is the case, how are the Hadamard-gates on the surface code implemented?

The physical entangling gate implemented experimentally is always a CZ gate, although we discuss logical entangling operations in terms of the transversal logical CNOT. To reflect the physical implementation, the circuit in Fig. 3b is shown after adjacent Hadamards from decomposing transversal CNOTs have been canceled, resulting in the effective logical circuit containing CZs. Logical Hadamards on the surface code, beyond those implicit in state initialization and measurement, are therefore not required. In Figs. 4-6, the logical CZ gates are transversal for the [[7,1,3]] and [[15,1,3]] codes. For the [[16,6,4]] CZ is not transversal (as for the surface code), but the identical logical circuit can be rewritten as transversal CNOTs with compiled Hadamards.

Revision: to clarify the distinction between CZs and CNOTs, we amend the caption of Fig. 3b to read

"Quantum circuits for realizing transversal gates and lattice surgery

operations. Logical CZ gates are ~~realized as~~ constructed from transversal CNOTs ~~and~~ with adjacent Hadamards canceled."

3. The manuscript would benefit from a clearer presentation of the role and extent of postselection in the various protocols. It would greatly improve clarity if the authors could provide a summary, e.g. in form of a table, detailing the decoding strategy, the presence of postselection and its type (based on confidence, loss-information, non-matching stabilizer values), and how each of these factors contributes to the reported acceptance rate. A comment on why certain postselection types are employed, e.g. why is postselection based on the loss-information used for the color codes but not for the surface code, would help to follow and understand each protocol.

We thank the Referee for these suggestions.

Revision: We have tabulated the types of decoding and postselection used throughout this work in an additional Supplementary Information file, titled

Summary of decoders

To further clarify the choices of acceptance fractions, we have added the following to the Methods section *"Additional experiment and data analysis details"*

"We use a lookup table for decoding and plot all three 3D color code curves with an acceptance fraction of 46% and the 2D color code with 74%, corresponding to a rescaling by the number of physical qubits in the code. This slightly sharpens the plateaus, which are otherwise smoothened in the presence of logical Z errors."

and, to the main text, added a reference to the Methods where we discuss the use of acceptance fractions in Figure 6:

"On the logical-qubit level we observe the expected correlations between logical outcomes corresponding to successful algorithm evolution, with a decay with distance corresponding to an effective algorithmic error rate that improves with decreasing entropy (corresponding to stricter postselection, Methods ~~lower acceptance fraction~~)."

4. A comment on how fault tolerance is ensured for the various protocols would be a helpful addition. For example, how is the encoding circuit shown in Extended Data Fig. 10 made fault-tolerant (by e.g. a transversal copy-and-verification approach)? The section "Universality and synthesizing arbitrary unitaries" does not mention fault tolerance. The authors write that by leveraging transversal teleportation, it is "ensured that all physical errors are

removed." How are all physical errors left behind, when (single) errors can propagate through transversal gates (onto single qubits in other blocks)? How is fault tolerance ensured, when e.g. each noisily encoded 3D color code block initially includes a single error on a data qubit? A more detailed discussion of these aspects would improve confidence in the robustness of the shown protocol.

While the hypercube encoding circuit used in these experiments is not fault-tolerant, it could be made so using flagged protocols as suggested by the Referee. Such fault-tolerant state preparation of the [[7,1,3]] code was shown in Ref. 11, and Ref. 120 further discusses a general verification scheme for these hypercube encoding circuits, leveraging code automorphisms for strict fault-tolerance.

Revision: We have added the following to the Methods section *"Additional experiment and data analysis details"*

"The [[7,1,3]] and [[16,6,4]] codes in Fig. 6, as well as the [[15,1,3]] in Fig. 4, are members of the family of quantum Reed-Muller codes based on the hypercube encoding circuit illustrated in ED Fig. 10a (see also Supplementary Videos) [120]. For each code, a different pattern of local Y($\pi/2$) pulses is applied while the entangling gate structure is the same; for the 2D [[7,1,3]] code, the fourth layer of gates is turned off. Although the encoding circuit alone is not fault-tolerant, a verification protocol [Ref. 120] or ancilla flag qubits [Ref. 11] can be directly added for future experiments."

The fault-tolerance of the state preparation is independent of the entropy removal properties of the logical teleportation circuit, which is transversal and thereby intrinsically fault-tolerant. When we say that all physical errors are removed, we mean to say that all types of physical error are removed by this protocol. The CZ gates ensure that physical Z errors do not propagate and physical X errors propagate through at most one gate. Furthermore, the preparation of clean input states both resets the code stabilizers to be +1 and further resets atomic degrees of freedom such as atom temperature and atom loss. This bounded error propagation and native reset is what we mean by "all physical errors are removed". Furthermore, the case where single qubit errors can propagate through more than one gate is studied in ED Fig. 10c,d by replacing the CZ gates with CNOTs. There we again find that the physical error propagation is bounded due to the regular transversal measurements, provided error correlations are properly accounted for.

Revision:

"To ensure that all types of physical error are removed and that computation is kept at constant entropy, we again leverage transversal teleportation."

and in the Figure 5 caption:

"Logical teleportation is used to ensure all types of physical error are removed."

5. In Extended Data Fig. 3, the authors explain that the non-destructive measurements take > 10 ms, which is longer than a full round of QEC (4.45 ms, line 962). This seems to be a bottleneck for the execution of large scale quantum computations that rely on measurements. In the outlook and methods section (line 902 and following), the authors discuss possible experimental improvements for single- and two-qubit gates, but not measurements. It would be a valuable addition to the discussion if the authors could extend these parts and comment on how these slow measurements can be a limiting factor, or why the authors believe that they are not.

We thank the Referee for this comment. While for simplicity we often chose to use comfortable measurement parameters in this work, we expect readout and reset times of a few milliseconds to be achievable with optimization, as demonstrated in the repeated Rabi calibration (Fig. 5b).

Revision: We have added a discussion of readout speed to the Methods section "*Processor clock speed*":

"In the repeated Rabi calibration in Fig. 5b, we did optimize for speed and achieved a cycle time of 4 ms. Although the imaging here was global as a demonstration of fast calibration, we expect such speeds can also be achieved in a zoned manner. Destructive measurement can be faster than the qubit re-use approach employed here, but the absence of loss information degrades QEC performance (Fig. 2) and further increases required qubit reloading rates for continuous operation. While non-destructive readout is slower, this may or may not bottleneck operations depending on e.g. when the next non-Clifford operation occurs. Comparing these holistically in the context of a whole architecture is an important avenue of future research.

6. The ML decoder used in this work requires extensive training on a given code size, followed by fine-tuning fed by experimental data. Working with a larger distance code (beyond d=5 explored in this work) seems to require re-training the respective neural networks entirely, with the training times and required samples probably not scaling favorably. Can the authors comment on the scalability of this approach?

We thank the Referee for this question. Understanding the scalability of dif-

ferent ML decoders with respect to many relevant metrics, including inference time, decoder performance, memory requirements, and training complexity is an important open research direction, noted in the Outlook. In particular, the details depend on the precise architecture used. The fully-connected neural networks employed for the surface code required re-training for each code distance and specific circuit (e.g., when changing the structure of dynamical decoupling pulses). Recent works on attention-based architectures, such as in Ref. 32 and Ref. 51 (*manuscript in preparation*), enable more flexibility for modifying the circuit without re-training, although the samples required for initial training times still scale exponentially in code distance [Ref. 32].

Revision: At the end of the Methods section *"ML decoder - surface code"* we have added

"Although the ML decoder used here is not directly scalable to high distance codes – for example requiring re-training for each code distance and specific circuit, with the number of training samples growing exponentially – exploring different extensions of such neural network architectures for scalable decoding is an exciting direction of ongoing research [Ref. 32, Ref. 51, Ref. 96]."

7. A key achievement of this work is the implementation of error-corrected universal logical quantum gates and their exploration in deep logical circuits. Eventually, for logical algorithms to outperform their counterpart on physical qubits, it is important to compare all logical operations (in particular the gates, of course) with their physical counterparts. For the (transversal) Clifford operations, the logical operations seem to outperform the respective physical gates. How does the performance of, e.g., the up-to-depth 4, H-T teleported logical single qubit Bloch sphere dynamics shown in Figure 4 compare to the respective physical operations? Do they operate below break-even? If not, can the authors comments on or, better, numerically estimate which increased code distances or experimental parameter improvements are needed to reach such beneficial logical fidelities of the universal gate set over the physical operations?

We thank the Referee for this insightful question. It is intrinsic to universal processing and synthesis of analog angles with digital gates that the logical qubit error grows linearly as the circuit depth increases, whereas the non-protected physical qubits will not get linearly worse for different angles. As such, it is clearest to compare the fidelity of the individual generated T states to the single-qubit preparation fidelity. For the T states shown in Fig. 4c, we find a fidelity of 98.2(2)%, while for physical qubits typical SPAM fidelity is 99.4%. The protocol for logical T state generation used here can be applied to larger code distance to further suppress errors, for example replacing the physical qubits of the [[15,1,3]] code with surface codes.

Revision: we have added additional discussion in a new subsection of *"Methods of universality"*, titled

"Small-angle synthesis"

In addition, there are some minor aspects, which should be addressed or clarified by the authors.
1. In Fig. 4, the authors prepare a logical state $|+\rangle_L$ and analyze the logical X-projection as a function of the rotation angle ϕ. However, the shown logical X-projection for $\phi = 0$ is -1, whereas one would expect $+1$ for an unrotated $+_L$-state.

We agree.

Revision: we have updated the x-axis of Fig. 4a and ED Fig. 9a to range from $\phi = -180$ to 180 degrees rather than 0 to 360 degrees, corresponding to applying an in-software Pauli Z_L.

2. It would be helpful if the authors could add the fidelities or logical error rates (and how they are calculated) for the synthesized states presented in Fig. 4. Fig. 4c shows the angular deviation from the ideal target rotation, but this alone does not provide a direct measure of the logical fidelity or error rate associated with the prepared states.

We thank the Referee for this suggestion.

Revision: We have added a new subsection ***"Small-angle synthesis"*** with these details in the Methods section *"Methods of universality"*.

3. In line 1219, the authors write "The curves are further normalized to highlight key features, with unnormalized data shown in ED Figure 9a." What kind of normalization is done?

We agree this is unclear in the current presentation. The logical expectation value is normalized by the purity and the stabilizers are normalized by the maximum expectation value.

Revision: In Line 1219, we have clarified the normalization procedure

"To highlight key features, the curves are further normalized ~~to highlight key features~~ by the purity (top) and maximum stabilizer

expectation value (bottom), with unnormalized data shown in ED Figure 9a."

We have further updated the caption of Fig. 4a to read

"3D Reed-Muller codes, only with all positive stabilizer signs, have an additional plateau at 45-degrees corresponding to non-Clifford T gates. Curves are normalized to unity maximum expectation (Methods)."

4. In the Methods Section (line 793), the authors write about the implemented lattice surgery protocol "A shot is counted as an error $Z_L^1 Z_L^2 = -1$ [...]". For an input state ++, the ZZ-measurement corresponds to a random projection and does not in itself imply an error. Is this additional postselection that is done and, if so, is it included in the stated acceptance rates?

We do not apply any postselection based on the measured ZZ value. The lattice surgery protocol is successful if the measured Pauli product $Z_L^1 Z_L^2$ from the ancilla measurement and from the transversal data qubit measurement agree (either both being +1 or both -1). We agree this is unclear in the original notation, which should instead state the product $(Z_L^1 Z_L^2)_{ancilla} \cdot (Z_L^1 Z_L^2)_{data} = -1$ corresponds to an error.

Revision: We have modified line 793 to read

"A shot is counted as an error~~, $Z_L^1 Z_L^2 = -1$,~~ if these two decoding procedures disagree."

The listed clarifications are required to improve the understandability of certain aspects. Overall, as outlined above, this work represents an impressive leap forward in the realization of fault-tolerant quantum computing. Therefore, we clearly recommend this manuscript for publication in Nature, provided the aspects outlined above are satisfactorily addressed.

References

[1] Rajeev Acharya, Dmitry A Abanin, Laleh Aghababaie-Beni, Igor Aleiner, Trond I Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Nikita Astrakhantsev, et al. Quantum error correction below the surface code threshold. Nature, 2024. doi:10.1038/ s41586-024-08449-y.

[2] Nathan Lacroix, Alexandre Bourassa, Francisco JH Heras, Lei M Zhang, Johannes Bausch, Andrew W Senior, Thomas Edlich, Noah Shutty, Volodymyr Sivak, Andreas Bengtsson, et al. Scaling and logic in the color code on a superconducting quantum processor. arXiv preprint arXiv:2412.14256, 2024.

doi:10.48550/arXiv.2412.14256.
[3] C Ryan-Anderson, NC Brown, CH Baldwin, JM Dreiling, C Foltz, JP Gaebler, TM Gatterman, N Hewitt, C Holliman, CV Horst, et al. High-fidelity teleportation of a logical qubit using transversal gates and lattice surgery. Science, 385(6715):1327–1331, 2024. doi:10.1126/science.adp6016.
[4] Ivan Pogorelov, Friederike Butt, Lukas Postler, Christian D Marciniak, Philipp Schindler, Markus Mueller, and Thomas Monz. Experimental fault-tolerant code switching. Nat. Phys., 2025

We thank the Referee again for their positive evaluation of our manuscript and helpful comments.

Reviewer: 3

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Reviewer: 4

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Reviewer: 5

Referee Report for "Architectural Mechanisms of a universal fault-tolerant quantum computer" by Bluvstein et al.

Summary and Recommendation

The authors demonstrate an integrated neutral-atom architecture that combines a large number of primitives and architectural elements for fault-tolerant QEC: repeated surface-code QEC with loss-aware decoding, logical two-qubit operations via both transversal coupling and lattice surgery, universal single-qubit gate synthesis through teleportation, mid-circuit measurement, qubit re-use, and active entropy removal. Overall, the work is very impressive and largely delivers on its promise to "implement all key elements of a universal, fault-tolerant quantum processing architecture."

My overall recommendation is publication after some minor revision. The results are highly significant and relevant. My only suggested revisions are to add some clarification and, in a few places, some qualification about the claims.

Key strengths

1. The authors show clear "per round" below-threshold evidence with quantified gains from erasure signals. At d=5, they show a nearly 2x lower error rate per round than at d=3 using a four-round memory experiment. Crucial to these improvments was the improved decoding using machine learning and information about loss to augment the bare symdrome. I appreciated the error budget figure as well (2f in the Supplement), which shows that 3/4 of the total error is split between SPAM and CZ gates, with about half of that due to loss. The error budget breakdown certainly suggests that there are substantial gains to be had by addressing this information at the code level, and I'm very pleased to see the large performance gains the authors have made in this area. While this is not quite the most stringent standard for quantifying QEC logical performance (as the authors discuss), it is nonetheless impressive.

2. The authors consider a comparison between transversal gates and lattice surgery to see which is more practical. They quantify sensitivity to ancilla measurement errors and demonstrate that this may lead to inherent limitations with lattice surgery approaches (at least in some regimes). They then identify an optimum of $\sim$3 transversal CNOTs per QEC round, matching a simple entropy-based model.

3. The authors demonstrate universal single-qubit logical gates via teleportation via the Reed-Muller code. This code has a transversal logical T gate, which the authors use to demonstrate universal logical gate synthesis, at least when the logical gates are transversal. They show how repeated teleportations generate increasingly dense images in the Bloch sphere when synthesized gates are interleaved with Hadamard rotations.

4. The authors tout the engineering advances that have improved mid-circuit measurement and re-use, for example the spin-to-position readout. It is not within my expertise to judge the difficulty of integrating these advances into the architecture, but these are clearly important advances for fault-tolerant operation.

5. The authors run deep circuits with constant internal entropy, which one

can interpret as saying they are correcting errors at about the same rate as the errors are generated. Such a steady state is characteristic of operation in a fault-tolerant regime. They have also run repeated encoded cluster-state computations and shown constant stabilizer error and rapid decay of physical-error correlations.

We thank the Referee for their positive evaluation of our manuscript.

Suggested revisions

1. I think it would be valuable to clarify the "threshold" language and long-round scaling. I found lines 120–129 lacking in sufficient detail in the main text, and potentially misleading. While the per-round result is strong, the text notes that the effective threshold can worsen by 1.15x with many rounds and 1.1x with 1 transversal gate per round. It would be valuable for the reader to include an explicit definition of "below threshold" used here, and (ii) scaling estimates or additional rounds (experiment or validated sim) showing whether the d-scaling persists when the number of rounds grows and when logic is interleaved. Finally, as the authors well know, there is a distinction between per-round improvements and per d rounds in a QEC cycle. I think it is a little late to first highlight this distinction in line 163, as then line 121 gives a misleading impression for what "below threshold" behavior means.

We thank the Referee for the suggested clarifications. As the Referee notes, below-threshold can be context-dependent and, in particular, depend on the specific circuit. To explicitly define that below-threshold for a particular circuit in this context refers to improving performance with increasing code distance, we have added the following:

Revision:

"Theoretically, logical errors can be exponentially suppressed as $(p/p_{\mathbf{th}})^{(d+1)/2}$ by increasing code distance d if the physical error rate p is below a characteristic threshold $p_{\mathbf{th}}$, i.e. below-threshold performance."

With respect to point (ii), we refer the Referee to detailed discussion and simulations of the threshold scaling, both in the limit of many rounds and with the inclusion of transversal gates, which is given in the Methods and plotted in Extended Data Figure 8. These show that the d-scaling does persist. We find in ED Fig. 8 that for our experimental error model and specific measurements that, in fact, we expect a negligible change of threshold, but by considering other error models we conservatively conclude that the measured threshold result can change up to 10% with logic gates and 15% with the limit of many rounds.

Revision: We have added a reference to the simulations in the ED Fig. 8 to the relevant part of the main text

"Note that while our observations show below-threshold behavior for repeated QEC rounds, theoretically the threshold can get worse by ≈ 1.15x in the limit of many repeated rounds and by ≈ 1.1x when incorporating ≈ 1 transversal gate per QEC round (see Methods and ED Fig. 8)."

We further agree with the Referee that there is a distinction between per-round improvement and per-d-rounds improvement: for the latter, one has to also account for the increase in number of total rounds for larger code distance, and roughly can change the relative performance of $r_{3/5}$ by a factor of 3/5. However, we emphasize that in our transversal approach, larger code distances do not require more rounds of QEC, and in fact the conventional d rounds factor is removed for realizing error-corrected algorithms. This is the subject of theoretical work in Refs. 12, 41, 44 and is also highlighted by our results in Figure 3. Therefore, the correct comparison for below threshold here is for a fixed number of rounds as a function of code distance.
As these points are all clearly subtle, and we agree the existence of subtleties should be stated clearly in line 121, we have added the following revision:

"We find that d=5 has a 2.14(13)x lower error per round than d=3, indicating below-threshold behavior for this four-round characterization circuit (Fig. 2d) (see also discussion below)."

2. To help readers interpret Fig. 4, a brief note on the T-state generation/verification procedure with representative figures of merit (e.g., typical fidelity and success probability) would be useful. A concise sentence outlining the dominant error sources across the teleportation path and an order-of-magnitude indication of the space–time cost versus target rotation precision (even as a back-of-the-envelope estimate or pointer to Methods/SI) would nicely connect the demonstration to algorithmic overheads.

We thank the Referee for this suggestion.

Revision: We have added fidelities and acceptance fractions for the small-angle states generated in Fig. 4 to the Methods in a subsection under *"Methods of universality"*, titled

"Small-angle synthesis"

We have further added a reference to this in the main text (while addressing the editor's request to not lengthen our manuscript)

"The resulting logical states span a range of points on the Bloch sphere and match the expected angles with high precision (Methods)~~(consistent with statistical uncertainty)~~."

3. Since several claims use varying acceptance fractions (AF), a brief clarification of how AF influences results, particularly the reported "constant entropy" behavior, would strengthen the claim. I am a little nervous considering entropy in the context of postselection, since I think this could hide many potential distortions. One suggestion: A compact comparison at one fixed AF together with a sentence on reservoir-depletion limits and how continuous reloading would alter achievable depth would make the interpretation straightforward. But really any additional discussion on AF in the context of the constant entropy claims (in particular) are suggested.

We thank the Referee for this insightful comment. In Fig. 6b, we find that the stabilizer expectation remains constant as a function of cycle. Additionally, in Fig. 6d we plot the 2D logical cluster state stabilizers at a fixed global acceptance fraction, again indicating constant entropy operation up until the atom reservoir depletes and lost atoms increase the entropy. This effect can be removed with continuous reloading, in which case we expect the constant entropy behavior to persist to arbitrary depths. To not be biased by the loss-induced entropy increase in the remainder of Fig. 6, we truncate to calculated logical correlations to early time layers (described in Methods).

Revision: We have modified the caption of Fig. 6d (while minimizing additional word count) to read

"Logical correlations persist while stabilizer error correlations rapidly decay. Depth of constant entropy operation is set by the reservoir depletion; continuous reloading can remove this constraint. See Methods for tabulation of postselection applied here and below. ~~Here and below, various degrees of acceptance fraction are applied (see Methods for tabulation of all details)~~."

Furthermore, in Fig. 6c,i, the lightest curves correspond to an acceptance fraction of 1 which show the characteristic decay of the logical correlations with distance as the logical operator propagates. This same behaviour is seen in the curves at lower acceptance fraction, where the postselection extends the correlation length due to the decreased entropy. As the Referee suggests, when applied incorrectly postselection can modify the effective entropy of the processor. In the study of logical correlations in Fig. 6, we use postselection based on a decoder confidence threshold, rather than accepting a fixed number of shots. This corresponds to postselection at constant entropy, in contrast to the shot acceptance fraction which artificially increases the entropy of high-weight oper-

ators (with higher effective error rate at the same acceptance fraction) relative to low-weight ones.

Revision: To emphasize this distinction, we have added additional discussion to the Methods section *Additional experiment and data analysis details*,

"In Figs. 6c,g we instead use a global confidence threshold for each curve, which is then converted to a mean acceptance fraction. In this way, each curve corresponds to a constant effective error rate (equivalently, constant effective entropy) for the logical operator independent of its weight, resulting in a reduced acceptance fraction for higher weight operators. In contrast, fixed-acceptance postselection would bias high-weight operators to a higher entropy relative to low-weight ones."

and a reference to this in the main text

"On the logical-qubit level we observe the expected correlations between logical outcomes corresponding to successful algorithm evolution, with a decay with distance corresponding to an effective algorithmic error rate that improves with decreasing entropy (corresponding to stricter postselection, Methods ~~lower acceptance fraction~~)."

Minor comments

1. Some phrases are a little awkward. Specifically "exponentially low, digitized errors" should maybe say low error rate. (Errors aren't "low".) Referring to generic 1Q gates as "analog-like" is a little misleading, especially given that they are implemented logically using digital gates. I appreciate what the authors are intending here, but I object to the term analog in this context, as I think it is misleading. The word analog in a similar context is repeated on line 203, and I think it is again being used incorrectly.

Revision: We have modified the text to read

"In this approach, entanglement between many physical qubits is used to encode error-corrected 'logical' qubits that can have exponentially low~~, digitized errors~~ error rate while performing arbitrary ~~analog-like~~ computations."

and

"These observations demonstrate that teleportation can be used as a powerful tool for universal processing, allowing precise ~~analog~~ small-angle rotations to be built from digital gates."

2. Figure 4a should mention explicitly in the caption that the traces are normalized. It is written in the axis, but it's not clear what "Norm." means, and this should be stated in the caption.

Revision: As proposed by the Referee, we have updated the caption of Fig. 4a to read

"3D Reed-Muller codes, only with all positive stabilizer signs, have an additional plateau at 45-degrees corresponding to non-Clifford T gates. Curves are normalized to unity maximum expectation (Methods)."

and have updated the Methods section "Additional experiment and data analysis details" accordingly

"To highlight key features, the curves are further normalized ~~to highlight key features~~ by the purity (top) and maximum stabilizer expectation value (bottom), with unnormalized data shown in ED Figure 9a."

3. Line 61 should perhaps say "polylogarithmic overhead" for precision.

We thank the Referee for highlighting this.

Revision: We have updated the text in Line 61 to read:

"... altogether achieving all ingredients for scalable (i.e., polylogarithmic overhead) realization of arbitrary quantum algorithms."

We have further updated the same phrasing in the following places:

"We then investigate logical entanglement using transversal gates and lattice surgery, and extend it to universal logic through transversal teleportation with 3D [[15,1,3]] codes, enabling arbitrary-angle synthesis with polylogarithmic overhead."

"This enables precise synthesis of rotation angles with a polylogarithmic number of steps (Fig. 4d inset), as expected by the Solovay-Kitaev theorem."

4. You already list some suboptimal choices affecting fidelity (e.g., AOD intermodulation, SLM inhomogeneity, beam profiles, magnetic-field noise). A brief "expected near-term improvements" paragraph quantifying the projected

logical-error gains from each item would be useful.

We agree with the suggestion.

Revision: At the end of the Methods section *"Details of processor configuration"* we have added

"These imperfections limited these deep-circuit measurements in particular, and by fixing them we can then recover nominal performance, corresponding to operating at $\approx 2\times$ below threshold as measured in Figure 2. The Methods section "Error budget and path to 10x below threshold" describes expectations on how we can improve this nominal performance further to scales approaching 5-10x below threshold."

Additional changes

We have made other minor changes, separate from the comments of the Referees:

1. As requested by the editor, we have shortened the abstract to 230 words with the change: **"Quantum error correction (QEC) is ~~believed to be~~ essential for the realization of large-scale quantum computers."**

2. In Extended Data Fig. 2b, the label "antinode" has been changed to "node".

3. We have slightly reworded a proof in the Methods section *"Logical entanglement and physical entanglement"* to clarify an assumption in the proof of the entanglement bound:
**"~~For the maximally mixed logical state ρ_L, the reduced density matrix on A has rank $2^{\min(k,d-1)}$, implying $S_{\rm PS} \geq \min(k, d-1)$. If $k \leq d-1$, then $S_{\rm PS} \geq k$, so $S_{\rm PS} \geq S_{\rm LO}$ always holds. If $k > d-1$, then $S_{\rm PS} \geq d-1$. Here, $S_{\rm LO} \leq S_{\rm PS}$ as long as $d-1 \geq \lfloor k/2 \rfloor$, i.e., $k < 2d$.~~ We consider such regions to remove state dependence. In stabilizer codes $S_{\rm PS}(A) = |A| - r_A$, where r_A counts stabilizers fully supported in A. We assume $r_A = 0$ for all $|A| \leq d-1$ (no stabilizer fully inside any correctable region); this holds, e.g., for every size-$(d-1)$ subregion in the $[[16,4,4]]$ code. Then $S_{\rm PS}(A) = |A|$, so for $|A| = d-1$ we have $S_{\rm LO} \leq S_{\rm PS}(A)$ whenever $\lfloor k/2 \rfloor \leq d-1$ (i.e., $k < 2d-1$).
Thus, for ~~any~~ a $[[n,k,d]]$ code with $k < 2d-1$ and $r_A = 0$ for all $|A| \leq d-1$ (e.g., $[[16,4,4]]$), the logical operator entanglement cannot exceed the physical entanglement available in any region of size $d-1$."**

Referee responses and summary of revisions 2

Nature Manuscript number: 2025-06-15475 Bluvstein

October 18, 2025

Reviewer: 1

I thank the authors for their thorough response to the questions and comments and all detailed revisions they made to the manuscript.

A few more comments regarding the responses to a-d:
(a) I agree that Steane-based QEC can work well provided a FT logical state preparation factory. Still, the word "robust" is slightly vague when discussing the fault-tolerance of a QEC protocol.
(b) The magic state distillation does start from an encoding circuit, but is not guaranteed to prepare an error-free target logical state (that is why magic state distillation is required). And for magic state distillation, FT error correction on magic states is required before creating the concatenated code from non-perfect magic states.
(c) I agree that one can achieve arbitrary angle rotation by switching between [[15, 1, 3]] and [[7, 1, 3]] codes. This is based on the (one-directional) transversal CNOT gate between the two codes though, which is specific to the choice of codes.
(d) I agree that transversal CNOT can perform logical teleportation, but then single-qubit errors would propagate through layers. The alternate approach requires transversal H, which is, again, specific to the choice of codes.

In general, I would still think this work is closer to "realizing several key ingredients of FTQEC using different, suitable codes" rather than "realizing FTQEC", so I thank the authors for softening their claim on the key aspects of this work. The code switching protocol between [[7, 1, 3]] and [[15, 1, 3]] can, in principle, realize FTQEC, based on these specific transversal gate gadgets: transversal CZ, one-directional transversal CNOT, transversal T in [[15, 1, 3]] and transversal H in [[7, 1, 3]]. This transversal gate set is so specific such that it is hard to imagine that for a general code switching protocol all those gates would work, and we have good reasons to believe that code switching between [[7, 1, 3]] and [[15, 1, 3]] is not the ultimate solution to practical FTQEC due to the low error threshold. Nevertheless, I can imagine "realizing FTQEC by switching between [[15, 1, 3]] and [[7, 1, 3]]" to be the next step of the authors' work, and this work indeed provides many, although not all, tools towards that goal. I therefore

recommend this work to be published in Nature.

We thank the Referee again for their positive evaluation of the manuscript, useful suggestions, and recommendation for publication. We agree that the specific realisation of FTQEC depends on the codes and circuit employed, and here we study one possible approach. Exploring different architectures remains an exciting ongoing research direction.

Reviewer: 2

The authors have carefully addressed my comments. Together with addressing the comments by the other reviewers, this has significantly improved clarity about key aspects of what has and what has not yet been achieved in this work. I am therefore happy to recommend publication of this very impressive advancement towards scalable fault-tolerant quantum computing.

I still have two comments I ask the authors to consider:
I thank the authors for the inclusion of the newly added Extended Data Fig. 8d, on the numerically simulated expected performance of the lattice surgery protocol for d=3 and d=5 for different numbers of stabilizer measurement rounds. This increases clarity regarding the choice of the number of rounds and of the relative performance for distance 3 and 5. Still, I think for clarity, it is important to highlight more explicitly that the surgery protocol with only 3 rounds (instead of 5) is not fault-tolerant in the strict sense as required for a fault-tolerant quantum circuit for such entangling operation via surgery on the d=5 code. I suggest to slightly rephrase the corresponding last sentence in the figure caption accordingly, to read e.g. "Whereas the protocol with 3 rounds for d=5 is not strictly fault-tolerant, we observe that the minimum error occurs at roughly 3 rounds for d=5, as opposed to 5."

We thank the Referee for the suggested clarification and have accordingly updated the Figure 3d caption.

Revision: **Numerical simulations for lattice surgery (ED Fig. 8d) show a minimum error-corrected fidelity at roughly 3 rounds for d=5. Note that error detection is used for lattice surgery in d (red marker) to further compensate for having $< d$ rounds (Methods).**

In the newly added subsection on "Small-angle synthesis", the sentence in line 1117 "By adding the appropriate feedforward at each teleportation step, this protocol becomes deterministic." is slightly ambiguous and could be misinterpreted as if the authors did this feedforward (experimentally). I understand that this is not what the authors want to express, but for clarity, to avoid such potential misconception, I suggest rephrasing the sentence slightly as "Adding the appropriate feedforward at each teleportation step would render the protocol deterministic."

We again thank the Referee for this suggestion and have updated the phrasing to remove potential misinterpretation.

Revision: **~~By~~ Adding the appropriate feedforward at each teleportation step would render this protocol ~~becomes~~ deterministic.**

Reviewer: 3

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Reviewer: 4

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Reviewer: 5

I have reviewed the authors' reply and their changes, and I am satisfied that the current article meets Nature's standards.

We thank the referee again for their positive evaluation of the manuscript and their useful suggestions.