nature portfolio

Peer Review File

Scaling and logic in the color code on a superconducting quantum processor

Corresponding Author: Dr Kevin Satzinger

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)
Scaling and logic in the color code on a superconducting quantum processor

Key results

The authors describe a demonstration the color code using a superconducting qubit processor.
They detail four experiments:
- The first demonstration of logical error rate suppression by increasing the code distance.
- Logical randomized benchmarking of single-qubit Clifford gates.
- State injection, including of magic states.
- State teleportation of a logical qubit using lattice surgery.

Originality and significance

The results presented are of immediate interest to many people in the quantum information community.
Together, the experiments present a compelling case for considering the color code as a candidate for building a large-scale fault-tolerant quantum computer.
I am sure that this work will spur much further research on this code family.

Data and methodology

The data and presentation are of excellent quality, with sufficient details on the circuits and the device.

Appropriate use of statistics

I find the use of statistics to be appropriate.

Conclusions

The conclusions and the interpretation of the data are robust.

Suggested improvements

I have some very minor suggestions for improvements:
- Start of Sec. II: I think it is confusing to conflate the two-dimensional and triangular color code, as the latter is a subclass of the former.
- End of Sec. V: I would recommend giving a definition or reference for the concept of the "Pareto front" as I don't think this will be familiar to all readers.
- End of App. C1: it was not clear to me how the two auxiliary qubits act as flags for each other. Could the authors provide more information, or a reference?
- Eq. (G1): possible typo brackets too small.

- App. I1: from Eq. (I7) onwards the "e" subscript of the entangelemnt fidelity disappears, which confused me at first reading.

References

N/A

Clarity and context

The abstract is clear and accessible and the introduction gives an appropriate overview.

Referee #2

(Remarks to the Author)
The authors implement color codes on a superconducting quantum computer. The authors show a suppression in logical error when increasing from d=3 to d=5 using a neural-network decoder. The authors prepare arbitrary states by state injection into a d=3 color code and teleportation between two d=3 color codes by lattice surgery. The authors also demonstrate logical randomized benchmarking showing that Clifford gates result in negligible error relative to simply performing error correction.

The authors implement the triangular version of the 6.6.6 hexagonal color code on a square lattice of qubits. Each color-code plaquette encodes a Z-type and X-type check and here these two checks are measured simultaneously using syndrome qubits encoded into a Bell state. This results in a cheap-fault-tolerant scheme relative to other flag methods and compatible with the underlying two-dimensional grid. The discussion of decoders in the appendix is illuminating and points to the benefits of neural-network decoders. The authors demonstrated state preparation for a range of magic states using single-qubit state preparation and Bell-state preparation.

The data is widely available and the methodology is clearly presented. The uncertainties presented in the final results are clearly described.

The conclusions follow from the data and analysis and are reliable. The forward-looking conclusion comparing the resources codes future surface codes and color codes is interesting but only supported by the calculations in the supplemental material. It is an important point to include and I understand why the calculations are secondary in this primarily experimental paper.

Areas for improvement:

In Fig. 5H and 5I, there is a large discrepancy between the measured Y values and the expected Y values. In the main text, the authors write "which we discuss in Appendix I 3." In the Appendix the authors suggest that has to do with an uncompensated virtual phase. It is not clear to this reviewer, if the authors applied the same virtual phase to the X and Y measurements or only to the X measurements. The authors are encouraged to calibrate the Y phase and then report the difference in the final additional phases required for the two measurements.

An error in Y measurements is also seen in the error corrected rotations in Figure 13 in the supplemental material. There it is attributed to Y being sensitive to "phase-flip and bit-flip" errors. This argument is not persuasive. Does the sensitivity of Y come from the decoder? If you trained it to assume Y and Z errors would logical X be further reduced?

In the main text in figure 4 c, is the in-phase oscillation of X and Y consistent with a small miscalibration in phi?

References: The references give appropriate credit to prior work. In Fig. 14, in the appendix it would be convenient to the reader to connect the colors with the appropriate references.

The paper is clearly written and easy to follow. The abstract captures the key results of the paper.

Referee #3

(Remarks to the Author)
The submitted manuscript, "Scaling and logic in the color code on a superconducting quantum processor," demonstrates several key aspects of fault-tolerant quantum computing using the color code on a 72-qubit superconducting processor. As a topological quantum code competing with the surface code, the two-dimensional color code offers advantages such as higher encoding rates and efficient logical operations. Previous works have demonstrated logical state memory and universal gate sets based on transversal gates and magic states in ion-trap systems. However, these works were limited to a code distance of 3 and did not reach the fault-tolerance threshold.

Specifically, this manuscript presents logical state memory (below threshold), benchmarking of logical single-qubit gates, magic state injection, and lattice surgery within the color code framework. Each of these components represents a significant result for the community, and, more importantly, this work is the first comprehensive experimental study of these aspects in the color code. The main contributions can be summarized as follows:

--Below-threshold logical state memory with color code. The manuscript demonstrates encoding of distance-3 and distance-5 color codes below the threshold, achieving a logical error suppression factor of 1.56. Essentially, this part follows a similar methodology to previous studies on surface codes (e.g., Nature volume 614, 676–681 (2023)), replacing the surface code with the color code. Notably, the numerical results suggest that if the physical error rate of the hardware were lower, the color code could surpass the surface code in performance, thereby reinforcing the significance of this work.

--Randomized benchmarking of logical single-qubit gates. The inherent properties of the color code facilitate the implementation of Clifford gates on single logical qubit. To the best of my knowledge, this is the first experimental demonstration of randomized benchmarking at the logical qubit level.

--Magic state injection and lattice surgery with color code. These are essential steps for achieving universal fault-tolerant quantum computation. Lattice surgery on color codes is demonstrated for the first time in a superconducting system, though some prior studies in ion-trap platforms have also explored these aspects. Additionally, the fidelity of the injected magic state reported is higher than the results from previous works.

In fact, these demonstrations cover nearly all essential components of fault-tolerant quantum computing using the color code (with the exception of magic state distillation). Overall, I find the manuscript suitable for publication in Nature but recommend addressing the following comments:

--The manuscript states that circuit transformations placing the measured qubits in lines reduce error rates. I have the following concerns regarding this claim: (a) This qubit arrangement is abandoned in the lattice surgery section. Does this imply that the circuit transformation approach is not broadly applicable, particularly in lattice surgery? (b) To what extent does this transformation affect the error rate? (c) If the circuit transformation method is not broadly applicable, could the design of superconducting chips optimized for color code encoding help circumvent this issue?

--The introduction mentions that the color code supports parallel Pauli operator measurements. However, this feature is not demonstrated in the lattice surgery section. Is there a fundamental difficulty preventing its implementation? This might be another crucial aspect in showcasing the advantage of the color code over the surface code.

--Throughout the manuscript, the Alpha Qubit decoder is primarily used. However, in the magic state injection section, the Möbius decoder is employed instead. What is the rationale behind this choice?

--The supplementary information compares the speed of different decoders, stating that MLE is about 1000 times slower than the Möbius decoder and 300 times slower than the neural-network decoder used in this work. However, is this comparison based on a code distance of 3 or 5? Clarifying the applicable scenario would improve the presentation of these results.

--In the magic state injection section, I noticed that the infidelity of the A_L state is significantly lower than that of the other two magic states. Additionally, in supplementary Fig. 13, the data point at ϕ=pi/4 for the A_L state is noticeably lower than the surrounding data points. Could the authors provide an explanation for this observation?

--The supplementary information states: “Note that we add a phase offset of 0.044 pi to compensate for a coherent phase rotation of the logical qubit, which we measured in a previous calibration measurement.” Could the authors elaborate on how the 7.9° offset was determined and the justification for this adjustment?

--The introduction claims that the color code can reduce space-time overhead by up to a factor of three. I believe this statement is misleading, as it implicitly assumes the same code distance for both codes. However, the logical error rates of the color code and surface code at the same code distance differ, particularly in lattice surgery.

--In Appendix H of the supplementary information, the first paragraph states: “use a fault-tolerant protocol to expand the color code...” This statement is incorrect because a protocol applied to a “d=1” code cannot be fault-tolerant.

We thank the referees for their insightful comments and suggestions. Based on their feedback, we feel we have improved the manuscript substantially. We note that we are under a length constraint, needing to reduce the main text length by 300 words in these revisions, so some content has to be addressed in the supplementary information.

Referee #1

Key results

The authors describe a demonstration the color code using a superconducting qubit processor.
They detail four experiments:
- The first demonstration of logical error rate suppression by increasing the code distance.
- Logical randomized benchmarking of single-qubit Clifford gates.
- State injection, including of magic states.
- State teleportation of a logical qubit using lattice surgery.

Originality and significance

The results presented are of immediate interest to many people in the quantum information community.
Together, the experiments present a compelling case for considering the color code as a candidate for building a large-scale fault-tolerant quantum computer.
I am sure that this work will spur much further research on this code family.

Data and methodology

The data and presentation are of excellent quality, with sufficient details on the circuits and the device.

Appropriate use of statistics

I find the use of statistics to be appropriate.

Conclusions

The conclusions and the interpretation of the data are robust.

We thank the referee for their clear summary of our work and their assessment of its strong research impact. We appreciate their positive comments on the technical content.

Suggested improvements

I have some very minor suggestions for improvements:
- Start of Sec. II: I think it is confusing to conflate the two-dimensional and triangular color code, as the latter is a subclass of the former.

We adapted the phrasing to be more clear on this point.

- End of Sec. V: I would recommend giving a definition or reference for the concept of the "Pareto front" as I don't think this will be familiar to all readers.

We edited this sentence to be more accessible.

- End of App. C1: it was not clear to me how the two auxiliary qubits act as flags for each other. Could the authors provide more information, or a reference?

We thank the reviewer for pointing out this confusing language. We edited that sentence to better convey the message. It now reads: "*In that way, the two auxiliary qubits detect errors on each other in addition to the errors on the data qubits [Gidney2023]*".

> - Eq. (G1): possible typo brackets too small.

We updated the formatting of the parentheses to improve readability.

> - App. I1: from Eq. (I7) onwards the "e" subscript of the entangelemnt fidelity disappears, which confused me at first reading.

We added the subscript on F_e in the rest of that section.

> ## References
>
> N/A
>
> ## Clarity and context
>
> The abstract is clear and accessible and the introduction gives an appropriate overview.

We appreciate all the referee's comments and have updated the manuscript based on each of their suggestions.

Referee #2

The authors implement color codes on a superconducting quantum computer. The authors show a suppression in logical error when increasing from d=3 to d=5 using a neural-network decoder. The authors prepare arbitrary states by state injection into a d=3 color code and teleportation between two d=3 color codes by lattice surgery. The authors also demonstrate logical randomized benchmarking showing that Clifford gates result in negligible error relative to simply performing error correction.

The authors implement the triangular version of the 6.6.6 hexagonal color code on a square lattice of qubits. Each color-code plaquette encodes a Z-type and X-type check and here these two checks are measured simultaneously using syndrome qubits encoded into a Bell state. This results in a cheap-fault-tolerant scheme relative to other flag methods and compatible with the underlying two-dimensional grid. The discussion of decoders in the appendix is illuminating and points to the benefits of neural-network decoders. The authors demonstrated state preparation for a range of magic states using single-qubit state preparation and Bell-state preparation.

The data is widely available and the methodology is clearly presented. The uncertainties presented in the final results are clearly described.

We appreciate the referee's clear description of our work and positive comments on the data and methodology.

The conclusions follow from the data and analysis and are reliable. The forward-looking conclusion comparing the resources codes future surface codes and color codes is interesting but only supported by the calculations in the supplemental material. It is an important point to include and I understand why the calculations are secondary in this primarily experimental paper.

We agree with the referee that the resource comparisons are an important topic, and we highlight in the conclusion that detailed resource estimation studies are important to establish a better comparison of surface code and color code computation. Due to length constraints and the experimental focus of the paper, it was necessary to keep the supporting calculations in the supplementary information.

Areas for improvement:

In Fig. 5H and 5I, there is a large discrepancy between the measured Y values and the expected Y values. In the main text, the authors write "which we discuss in Appendix I 3." In the Appendix the authors suggest that has to do with an uncompensated virtual phase. It is not clear to this reviewer, if the authors applied the same virtual phase to the X and Y measurements or only to the X measurements. The authors are encouraged to calibrate the Y phase and then report the difference in the final additional phases required for the two measurements.

We rephrased the explanation in Appendix I3 to be clearer on the fact that the virtual phases were not applied in the Y measurements, while they were applied in the X measurement (matched basis). It now reads:

In particular, in our matched basis experiments (i.e. experiments with a deterministic outcome such as measuring $|+_L\rangle$ in the X basis), we <u>calibrate and apply</u> virtual phases to compensate for the Z-rotations. However, <u>no such phases were calibrated or applied</u> in the non-deterministic mismatched basis circuits (e.g. measuring $|+_L\rangle$ in the Y basis), which may explain the observed bias.

Unfortunately, it is not possible to take any additional data on this device at this stage, but we intend to calibrate these phases even in unmatched basis circuits in future work, as suggested by the referee.

> An error in Y measurements is also seen in the error corrected rotations in Figure 13 in the supplemental material. There it is attributed to Y being sensitive to "phase-flip and bit-flip" errors. This argument is not persuasive. Does the sensitivity of Y come from the decoder? If you trained it to assume Y and Z errors would logical X be further reduced?

We agree with the reviewer that the decoder may also influence the ability to accurately correct Y errors. In the state injection experiments, we use the Möbius decoder with a device-agnostic SI1000 detector error model. This decoder does not explicitly model correlations between X- and Z-type syndromes, and as a result, may have limited ability to identify and correct Y errors effectively.

We expand our explanation to incorporate this aspect:

We observe the largest discrepancy between decoded and post-selected values for Y_L, which we attribute to both its sensitivity to both bit-flip and phase-flip errors, and the limited ability of the Möbius decoder to capture correlations between X and Z syndromes.

> In the main text in figure 4 c, is the in-phase oscillation of X and Y consistent with a small miscalibration in phi?

Yes, the small in-phase oscillation of the expectation value of X and Y is consistent with a small miscalibration in the azimuthal angle ϕ as a function of the polar angle. While we have not investigated this aspect in detail, we attribute it to the fact that we use an arbitrary angle single-qubit Y gate on the injection qubit obtained by scaling an already calibrated pi-pulse (DRAG pulse), and for these arbitrary rotations we do not calibrate for any potential phase error (unlike our standard pi and pi-half DRAG pulses which we otherwise use in our circuits). Note that this phase error can very easily be accounted for by adapting the subsequent virtual-Z rotation on the injection qubit; but we did not do this for the sweep in Fig. 4c, where we kept ϕ fixed at 0 for simplicity.

> References: The references give appropriate credit to prior work. In Fig. 14, in the appendix it would be convenient to the reader to connect the colors with the appropriate references.

We appreciate the suggestion on Fig. 14 and added the connection between colors and references in the caption.

> The paper is clearly written and easy to follow. The abstract captures the key results of the paper.

We thank the referee for all their comments, which have resulted in several important improvements.

Referee #3

> The submitted manuscript, "Scaling and logic in the color code on a superconducting quantum processor," demonstrates several key aspects of fault-tolerant quantum computing using the color code on a 72-qubit superconducting processor. As a topological quantum code competing with the surface code, the two-dimensional color code offers advantages such as higher encoding rates and efficient logical operations. Previous works have demonstrated logical state memory and universal gate sets based on transversal gates and magic states in ion-trap systems. However, these works were limited to a code distance of 3 and did not reach the fault-tolerance threshold.
>
> Specifically, this manuscript presents logical state memory (below threshold), benchmarking of logical single-qubit gates, magic state injection, and lattice surgery within the color code framework. Each of these components represents a significant result for the community, and, more importantly, this work is the first comprehensive experimental study of these aspects in the color code. The main contributions can be summarized as follows:

We thank the referee for their clear description of our work and its several significant contributions.

> --Below-threshold logical state memory with color code. The manuscript demonstrates encoding of distance-3 and distance-5 color codes below the threshold, achieving a logical error suppression factor of 1.56. Essentially, this part follows a similar methodology to previous studies on surface codes (e.g., Nature volume 614, 676–681 (2023)), replacing the surface code with the color code. Notably, the numerical results suggest that if the physical error rate of the hardware were lower, the color code could surpass the surface code in performance, thereby reinforcing the significance of this work.
>
> --Randomized benchmarking of logical single-qubit gates. The inherent properties of the color code facilitate the implementation of Clifford gates on single logical qubit. To the best of my knowledge, this is the first experimental demonstration of randomized benchmarking at the logical qubit level.
>
> --Magic state injection and lattice surgery with color code. These are essential steps for achieving universal fault-tolerant quantum computation. Lattice surgery on color codes is demonstrated for the first time in a superconducting system, though some prior studies in ion-trap platforms have also explored these aspects. Additionally, the fidelity of the injected magic state reported is higher than the results from previous works.

We appreciate the detailed description of these main contributions, placing them in context with previous research results.

> In fact, these demonstrations cover nearly all essential components of fault-tolerant quantum computing using the color code (with the exception of magic state distillation). Overall, I find the manuscript suitable for publication in Nature but recommend addressing the following comments:

We appreciate the referee's recommendation to publish in *Nature* and for their constructive comments, which we address below.

> --The manuscript states that circuit transformations placing the measured qubits in lines reduce error rates. I have the following concerns regarding this claim: (a) This qubit arrangement is abandoned in the lattice surgery section. Does this imply that the circuit transformation approach is not broadly applicable, particularly in lattice surgery? (b) To what extent does this transformation affect the error rate? (c) If the circuit transformation method is not broadly applicable, could the design of superconducting chips optimized for color code encoding help circumvent this issue?

We thank the reviewer for the thoughtful questions and address each point below.

(a) We believe the circuit transformations are also applicable to lattice surgery. However, due to the complexity of the lattice surgery experiment and our focus on demonstrating functionality rather than optimizing performance, we chose not to apply them in this context. Moreover, the lattice surgery circuit involved only four mid-circuit measurements, in contrast to the many repeated cycles in memory or logical benchmarking experiments, where such transformations have greater impact on the relevant metrics.

(b) While we did not conduct a controlled A/B test on the current device, we observed in a previous-generation processor (not included in the manuscript) that omitting circuit transformations resulted in a ~50% increase in logical error rate in a distance-three memory experiment. That said, the benefit of circuit transformations likely depends on the specific measurement-induced dephasing rate, which varies substantially across the device (most qubits exhibit little effect, but some qubits exhibit a stronger effect, which may be related to the frequency and coupling arrangements).

(c) The current chip was optimized for surface code experiments, where aligning auxiliary qubits along a common readout line occurs naturally. However, we agree that chip design and calibration improvements could remove the need for these circuit transformations. Possible strategies include increasing frequency separation between readout resonators on a shared readout line, using individual Purcell filters as demonstrated in Krinner et al., *Nature* (2022), and employing pulse shaping to reduce the spectral width of the readout pulse.

We summarize these insights in a new paragraph of the lattice-surgery appendix:

Our primary goal in this lattice surgery experiment is to demonstrate functionality, and given circuit complexity, we choose not to apply the circuit transformation introduced in Appendix C3. Nevertheless, we believe that such transformations can in principle still be applied in the context of lattice surgery. In future work, the impact of measurement-induced dephasing may be reduced through a combination of circuit-level techniques (such as these transformations) as well as hardware design choices, including wider frequency spacing of readout resonators, individual Purcell filters [Krinner2022], or pulse shaping.

> --The introduction mentions that the color code supports parallel Pauli operator measurements. However, this feature is not demonstrated in the lattice surgery section. Is there a fundamental difficulty preventing its implementation? This might be another crucial aspect in showcasing the advantage of the color code over the surface code.

The referee is correct: we do not experimentally test the parallel Pauli measurements. We do not believe there is any fundamental difficulty. We did not implement such measurement because it would require a third logical qubit that can connect via lattice surgery to one of the two others, and this did not fit on the 72-qubit processor on which this experiment was executed.

> --Throughout the manuscript, the Alpha Qubit decoder is primarily used. However, in the magic state injection section, the Möbius decoder is employed instead. What is the rationale behind this choice?

There is no fundamental reason why AlphaQubit could not be used for the injection experiment. However, when full post-selection is applied (i.e., when all runs with non-trivial syndromes are discarded), no syndrome information remains for the decoder to process. In such cases, AlphaQubit offers no advantage over the Möbius decoder, and the resulting fidelity is effectively independent of the decoder used. While the decoder can influence results under partial or no post-selection, we do not expect substantial differences in this state injection experiment on a d=3 logical qubit, given that AlphaQubit's advantage over Möbius is most pronounced at distance d=5 (see also Fig. 10). For convenience, we therefore opted to use the Möbius decoder for the injection experiment, which avoids the need to generate a large training dataset for AlphaQubit.

To clarify this aspect, we significantly expanded the state injection appendix to describe the post-selection schemes in more detail. Specifically, we also added a paragraph to clarify the influence of the decoder on the state injection:

For convenience, we use the Möbius decoder along with a device-agnostic SI1000 noise-model prior for all state-injection experiments. When full post-selection is applied, the achieved infidelity is independent of the decoder, as no syndrome information remains for the decoder to process. For partial or no post-selection, more elaborate decoders or priors could offer modest improvements in performance, though they ultimately would not reach the infidelity levels achieved with full post-selection.

> --The supplementary information compares the speed of different decoders, stating that MLE is about 1000 times slower than the Möbius decoder and 300 times slower than the neural-network decoder used in this work. However, is this comparison based on a code distance of 3 or 5? Clarifying the applicable scenario would improve the presentation of these results.

We thank the reviewer for the question. We now clarify the relative timings of the various decoders at distance-5 specifically. These numbers could change based on the number of detection events and optimization of each decoder for speed, but they should be representative enough to make order of magnitude comparison in the speed of the different decoders. This paragraph now reads:

First, [the MLE decoder] is considerably slower than the other decoders. Specifically, it is currently about 880 times slower than the Möbius decoder and 260 times slower than the neural-network decoder when applied to the experimental data of the distance-five color code. While none of the decoders used in this work were fully optimized for speed, such optimizations will be necessary in the future to enable real-time decoding.

> --In the magic state injection section, I noticed that the infidelity of the A_L state is significantly lower than that of the other two magic states. Additionally, in supplementary Fig. 13, the data point at ϕ=pi/4 for the A_L state is noticeably lower than the surrounding data points. Could the authors provide an explanation for this observation?

We thank the reviewer for this observation. As noted, the infidelity of the $|A_L\rangle$ state appears lower than that of the other two magic states when full post-selection is applied (rightmost data points in Fig. 4e). However, considering the 95% bootstrapped confidence intervals indicated by the error bars, the infidelities of all three states exhibit some overlap. Moreover, without post-selection, the infidelity of $|A_L\rangle$ is not lower than that of the other magic states. A similar trend is observed in Supplementary Fig. 13: the data point at $\phi=\pi/4$ appears lower for $|A_L\rangle$

compared to the surrounding data points, but this feature does not appear in the absence of post-selection.

At present, we do not have sufficient evidence to conclude that the infidelity of $|A_L\rangle$ is systematically lower. However, this could be an interesting feature to explore in future work.

We added the following sentence to clarify the matter:

In our realization, the infidelity of $|A_L\rangle$ appears lower than that of the other two magic states under full post-selection, but we note that the difference is not statistically significant within the 95 % bootstrapped confidence intervals (indicated by the error-bars).

> --The supplementary information states: "Note that we add a phase offset of 0.044 pi to compensate for a coherent phase rotation of the logical qubit, which we measured in a previous calibration measurement." Could the authors elaborate on how the 7.9° offset was determined and the justification for this adjustment?

Using logical state tomography, we observed a consistent Z-phase shift between the target and injected logical states. Since this phase shift is systematic, it is straightforward to correct it by adding a constant offset of -7.9° to the virtual-Z rotation of the injection qubit (i.e. instead of performing a rotation of $Z(\phi)$, we perform $Z(\phi - 7.9°)$. While we did not investigate the physical origin in detail in this context, flux crosstalk or settling could cause a coherent frequency shift on the injection qubit compatible with this observation.

We adapted the sentence to clarify that the phase offset is determined using logical tomography:

Note that we add a phase offset of 0.044π to compensate for a coherent phase rotation of the logical qubit determined using logical state tomography in a prior calibration measurement.

> --The introduction claims that the color code can reduce space-time overhead by up to a factor of three. I believe this statement is misleading, as it implicitly assumes the same code distance for both codes. However, the logical error rates of the color code and surface code at the same code distance differ, particularly in lattice surgery.

We agree that in practice the color code and surface code may require different code distances to reach a target logical error rate. To avoid any confusion, we rephrase the sentence to omit the overhead reduction factor as it depends on the physical error rates and other details of the implementation. It now reads:

Color codes also enable multi-qubit entangling operations with less space-time overhead thanks to their capability for simultaneous fault-tolerant multi-qubit Pauli measurements [Thomsen2022].

The cited study by Thomsen et al. (2022) includes a detailed circuit-noise model analysis comparing the two codes, accounting for differences in required distances. That analysis estimates a space-time overhead reduction for the color code by a factor between 1.1 and 2.5, depending on the physical error rates. As noted by the authors of the study, this comparison does not account for recent improvements in color code syndrome extraction circuits and decoders, which could further increase the improvement factor.

> --In Appendix H of the supplementary information, the first paragraph states: “use a fault-tolerant protocol to expand the color code...” This statement is incorrect because a protocol applied to a “d=1” code cannot be fault-tolerant.

We appreciate the reviewer’s precision here. We have updated the sentence to avoid the mention of fault-tolerance. The protocol itself can be used to expand color codes from an arbitrary distance d_1 to a distance $d_2 > d_1$ and is in general fault-tolerant if $d_1 \geq 3$. However, as pointed out by the reviewer, in the case of state injection we start with a “d = 1” code, which isn’t fault-tolerant.