Demonstration of logical qubits and repeated error correction with better-than-physical error rates

M. P. da Silva,[1] C. Ryan-Anderson,[2] J. M. Bello-Rivas,[1] A. Chernoguzov,[2] J. M. Dreiling,[2] C. Foltz,[2]
J. P. Gaebler,[2] T. M. Gatterman,[2] D. Hayes,[2] N. Hewitt,[2] J. Johansen,[2] D. Lucchetti,[2] M. Mills,[2] S. A. Moses,[2]
B. Neyenhuis,[2] A. Paz,[1] J. Pino,[2] P. Siegfried,[2] J. Strabley,[2] S. J. Wernli,[1] R. P. Stutz,[2] and K. M. Svore[1]

[1]*Microsoft Azure Quantum*
[2]*Quantinuum*
(Dated: April 2, 2024)

The promise of quantum computers hinges on the ability to scale to large system sizes, e.g., to run quantum computations consisting of more than 100 million operations fault-tolerantly. This in turn requires suppressing errors to levels inversely proportional to the size of the computation. As a step towards this ambitious goal, we present experiments on a trapped-ion QCCD processor where, through the use of fault-tolerant encoding and error correction, we are able to suppress logical error rates to levels below the physical error rates. In particular, we entangled logical qubit states encoded in the $[[7, 1, 3]]$ code with error rates $9.8\times$ to $500\times$ lower than at the physical level, and entangled logical qubit states encoded in a $[[12, 2, 4]]$ code with error rates $4.7\times$ to $800\times$ lower than at the physical level, depending on the judicious use of post-selection. Moreover, we demonstrate repeated error correction with the $[[12, 2, 4]]$ code, with logical error rates below physical circuit baselines corresponding to repeated CNOTs, and show evidence that the error rate per error correction cycle, which consists of over 100 physical CNOTs, approaches the error rate of two physical CNOTs. These results signify an important transition from noisy intermediate scale quantum computing to reliable quantum computing, and demonstrate advanced capabilities toward large-scale fault-tolerant quantum computing.

Quantum computers have the potential to solve important classically-intractable problems, however doing so requires improving error rates well beyond those of the underlying physical hardware. The development of quantum error correction and fault-tolerant quantum computing was a major theoretical breakthrough that paved the way for the implementation of reliable quantum computers [1–7]. Without quantum fault-tolerance, there is little to no indication that quantum computers can solve important practical problems that are outside the reach of modern day supercomputers and machine learning [8, 9]. The experimental challenges remain significant, as fault tolerance requires that physical error rates be sufficiently low before the overhead of error correction leads to an improvement over physical, non-fault-tolerant operations [3, 5–7, 10], but steady progress has been made across several different platforms. In fact, several experiments have shown indications of physical error rates approaching this important so-called threshold [11–14], while others have demonstrated operations on multiple logical qubits [15–17]. However, to the best of our knowledge none of these experiments have demonstrated logical error rates better than the physical error rates— a notable exception being a demonstration of Bell correlations that are stronger than physical correlations [18].

Our goal is to demonstrate the transition from noisy intermediate scale quantum computing to reliable quantum computing [19, 20], through the co-optimization of hardware and software with a present-day commercial quantum processor. Namely, we aim (1) to show convincingly a large separation between logical and physical error rates, (2) in a setting where *all single circuit faults are corrected*, while (3) using logical circuits representative of what would be used for computation, e.g., the preparation or use of logical entanglement.

To this end, we demonstrate several fault-tolerant protocols in a commercial trapped-ion quantum charge-coupled device (QCCD) processor [21] and show that the observed logical error rates are conclusively lower than the error rates for their (unencoded) physical counterparts.

I. METHODOLOGY

Our approach largely builds on Gottesman's proposal in Ref. 22. Namely, we benchmark *complete quantum circuits* [23], and contrast the error rates of the classical outputs of these circuits—a comparison between the outputs of the unencoded physical circuit to that of the corresponding fault-tolerantly encoded circuit on the same hardware. We deviate from Gottesman's proposal in two ways – we consider a different metric for the comparison of quantum circuits, and we allow for slightly more general state preparations.

The metric considered in Gottesman's proposal is the *total variation distance* between the output distribution of the ideal circuit and the experimental circuits (encoded or otherwise). For our proposal we also consider the statistical distance between outputs, but we add classical processing of the measurement outputs so that it is possible to determine success or failure for each individual run of the experiment. For example, when preparing the state $\frac{|00\rangle}{\sqrt{2}} + \frac{|11\rangle}{\sqrt{2}}$ and measuring both qubits in either the Z or the X basis at random, we consider success if both outcomes agree and failure if they disagree. Due to the ad-

ditional classical processing (i.e., comparing the two bits and reporting only their parity) this metric is generally weaker than the distance proposed by Gottesman, but it simplifies the analysis and yields estimates with much lower uncertainty (an important consideration in an experimental setting with finite resources). Moreover, in many cases of interest, experiments of this type allow for direct estimates of the state fidelity through the measurement of the parity of the stabilizer group elements [24, 25].

We also deviate from Gottesman's proposal by allowing that the qubits be prepared in some finite set of fixed states instead of only the $|0\rangle$ state, since it is natural to consider specialized preparations of resource states in the encoded setting. In the unencoded setting the preparation will reduce to preparation of qubits in the $|0\rangle$ state and application of various gates.

II. HARDWARE PLATFORM

All reported demonstrations here were performed on Quantinuum's H2 trapped-ion processor. The device was recently reported on in detail in Ref. 21, and we give a brief overview here.

H2 is a shuttling-based trapped-ion QCCD device [26, 27]. H2 has all the necessary ingredients for state-of-the-art quantum error correction experiments: high-fidelity state-preparation and measurement (SPAM) with an error rate of 0.15%, two-qubit gates with an error rate of 0.14(1)%, long range connectivity (a key enabler for running some of the circuits described here), and mid-circuit measurement and reset with crosstalk errors $\leq 2 \times 10^{-5}$.

The H2 processor is commercially available through both Quantinuum [28] and Azure Quantum [29]. All experiments in Sections III B and IV were submitted through the Azure Quantum software stack using the Quantum Intermediate Representation (QIR) [30]. For the experiments described in Section IV, additional compiler customization were applied, as described in that section in more detail.

III. IMPROVED LOGICAL ENTANGLEMENT

Entanglement is a quantum hallmark, and over the last 20 years the preparation and measurement of Bell states has been a baseline demonstration for credible physical implementations of quantum computers. It is natural to consider the same baseline in the context of logical qubits [31], as the complexity of the fault-tolerant preparation already becomes apparent. Several recent experiments have demonstrated the preparation of logical Bell-pairs in quantum error correction codes [16–18], and although these results are remarkable in their own right, only one of them [18] has demonstrated a logical error rate modestly better than physical error rate. Here we describe how much lower logical error rates can be obtained using the [[7, 1, 3]] Steane code and a [[12, 2, 4]] code.

A. Steane code

We first present results using the [[7, 1, 3]] Steane code [2] to prepare a high-fidelity Bell state. The Steane code, or distance-three color code, has been used in several demonstrations of logically encoded circuits [12, 16–18, 32, 33], partially because of its relatively low space-time overhead, its simple preparation and measurement protocols due to its CSS nature, and all single and two qubit Clifford gates being transversal for the code.

1. Circuits

The circuit components used to generate a high-fidelity Bell state were previously demonstrated in Refs. 12 and 18. The logical program to prepare a logical Bell resource state using the Steane code is in Fig. 1. The preparation includes encoding circuits to initialize two logical qubits to $|0\rangle$, transversal single and two-qubit Clifford gates, flagged syndrome extraction, and destructive logical measurements. Each logical qubit has seven data qubits and three ancilla qubits, leading to the experiments having a total of 20 physical qubits.

The encoding circuit is made fault-tolerant using a scheme by Goto [34] which involves a non-fault tolerant encoding circuit of $|0\rangle$ followed by a verification step where an ancilla qubit measures the logical Z operator. Upon failure to verify the preparation a logical $|0\rangle$ state, the qubits can be conditionally reset and the fault-tolerant preparation can be re-attempted in a repeat-until-success fashion (as seen in Ref. 12) or pre-selected upon verification. For these experiments, we chose to repeat the preparation of the $|0\rangle$ up to three times. Once the $|0\rangle$ states are verified, the Goto scheme ensures that state preparation results in at most weight-one faults due to any single faulty gate.

After both logical qubits are prepared in $|0\rangle$, we apply a transversal Hadamard to one of the logical qubits, followed by a transversal CNOT between the two logical qubits. Ideally, this circuit would produce $(|00\rangle + |11\rangle)/\sqrt{2}$.

We attempt to verify this by following the CNOT with one round of flagged syndrome extraction on each logical qubit based on a scheme by Chao and Reichardt [35] in which three syndromes are measured in parallel. The main difference compared to the original scheme is that we do not follow the flagged syndrome measurements with a conditional set of unflagged syndrome measurements. This is because we are treating the Bell state as a fixed resource state independent of the computation, and much like the the logical $|0\rangle$, pre-selection of such resources has no negative impact on scalability [1, 36, 37].

After running syndrome extraction, logical single-qubit transversal gates are applied to measure in appropriate logical Pauli bases, and then we destructively measure the data qubits. We measure both qubits in the logical X, Y, and Z bases, which can be performed transversally in the Steane code. The destructive measurements

of data qubits not only allows one to determine logical outcome but can also be used to determine syndromes. At the end of each destructive logical measurement, these syndromes are used to generate a correction to the logical outcome using a lookup table decoder. These corrected outcomes are determined by running the decoder during the hybrid quantum/classical program on the device and not determined afterwards. The lookup table is relatively simple and the same one used in Ref. 12. Syndromes are decoded independently for each logical measurements. Further improvements could be made by incorporating experimental bias noise in the construction of the decoder, as well as decoding over all syndromes generated by both logical qubits [17, 38–40].

The $[[12, 2, 4]]$ code also studied in this work does not admit for a transversal fault-tolerant Y eigenbasis measurement, so to facilitate comparison we focus on the error rate for Bell experiments where we measure only X and Z parities, and denote it E_{xz}. Simply put, E_{xz} is defined as the number of "incorrect" measurement results divided by the total number of trials. Only measuring in the X and Z basis is related to measuring the stabilizers of the target state. These measurements can provide a bound on the fidelity of the operation, but since we can also measure in the Y basis when using the Steane code, we also report estimates of the process fidelity in Appendix C. The derived process fidelities give a more conservative estimate of the overall performance of the Bell preparation protocol; however, the process fidelities paint a similar picture to the estimates given by E_{xz}, where the encoded circuits are statistically significantly out performing the results of the unencoded circuit analog.

Besides evaluating E_{xz} in the presence of quantum error correction, we also re-analyze the corrected outcomes calculated by the device by excluding any results for which non-trivial syndromes were recorded during the destructive measurement phase and the evaluate the Steane code as a quantum error detecting (QEC) code. This mode of operation does not satisfy the requirement that all single faults must be corrected, but allows us to evaluate the probability of the environment applying pure logical errors (thus serving as an upper bound on QEC performance) as well as evaluate the rejection rate the cost of using the Steane code in a QEC manner and whether that cost is worth paying to further suppress noise.

As described in Section I, we compare E_{xz} of physical (unencoded) circuits and logical circuits. The physical circuits are identical to what is described in Fig. 1, with the exception that no syndrome extraction, pre-selection, or post-selection is applied. Since H2 has four gate zones, four copies of the physical circuits were executed in parallel for each run, totaling eight physical qubits per program.

2. Experimental results

The experimental results for the both the Steane code and physical level Bell state preparation are summarized
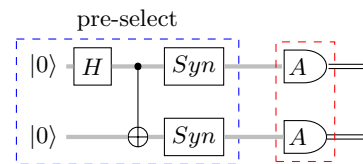

FIG. 1. High-level depiction of the logical program of the Bell resource-state preparation using the Steane code. The blue dashed box indicates the pre-selected portion of the circuit where both the verification of $|0\rangle$ and trivial measurement results of syndrome extraction rounds (the boxes labeled Syn) are used to verify the creation of a resource state. The experiments are analyzed using error correction (with pre-selection) or error detection (with pre- and post-selection) independently. Post-selection accepts experiments where the syndrome inferred from the logical measurements (red dashed boxes) is trivial.

in Table I and Fig. 2 (see Appendix A for details of the statistical analysis and Appendix C for additional data).

We ran a total of $411,600$ unencoded experiments (four Bell state preparation and measurement circuits per program). When restricted to X and Z basis experiments, $274,400$ experiments were ran at the unencoded level, respectively. For both the unencoded and encoded set of experiments, we ran an equal number of sub-experiments measuring X, Y, and Z correlations. Of the $411,600$ physical experiments, $1,897$ measured the incorrect parity. In particular, the wrong parity was measured 572 out of $137,200$ times for X parity, 530 out of $137,200$ times for Y parity, and 795 out of $137,200$ times for Z parity. This results in an error rate $E_{xz} = 0.50\%^{+0.03\%}_{-0.03\%}$. Looking at the outcomes for separate bases (see Table IV), it is apparent that the Z measurements experience an increased error rate of approximately 0.6% compared to about 0.4% for the X and Y. This difference may arise due to biased noised in the two-qubit gates.

For the logical experiments with error correction and pre-selection, $12,100$ experiments were ran for each of X, Y, and Z correlations, for a total of $36,300$ experiments. About $9,000$ shots were pre-accepted for each basis resulting in a pre-acceptance rate of about 75%. Out of the $9,025$ pre-accepted X correlations experiments, 337 non-trivial syndromes were measured and 9 experiments resulted in the wrong parity measurement. For the $9,082$ pre-accepted Y correlation experiments, 417 non-trivial syndromes were measured and 8 incorrect parity measurements were made. For Z correlations, of the $9,010$ pre-accepted experiments, 309 had non-trivial syndromes and 0 measurements resulted in measuring the wrong parity.

Post-selecting on non-trivial syndromes (so no error correction is performed, only error detection), the total acceptance rate goes from $\approx 75\%$ to $\approx 72\%$, and 0 incorrect parity outcomes are observed for the X and Z correlations out of $8,688$ and $8,701$ post-accepted experiments, respectively. For the Y, only 2 measurements out of $8,665$ experiments accepted in post-selection (post-
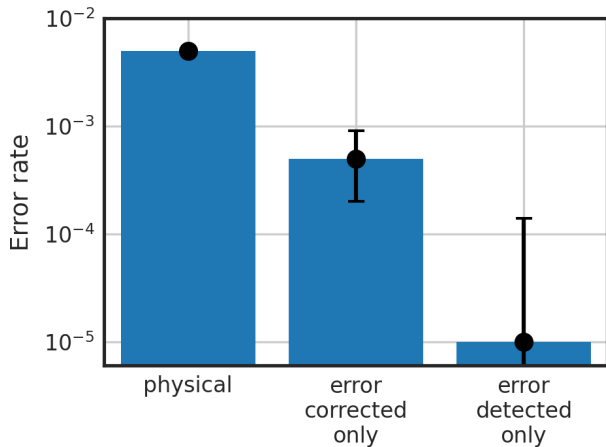
FIG. 2. Results for the Bell state preparation in the $[[7,1,3]]$ Steane code comparing physical level, error correction, and error detection experiments. There is a statistically significant separation between the physical and encoded results given the errors bars, which indicate 95% confidence intervals.

accepted) resulted in a measurement of the wrong parity.

Based on these observations, the error rates E_{xz} were determined to be $0.50\%^{+0.03\%}_{-0.03\%}$, $0.05\%^{+0.04\%}_{-0.03\%}$, and $0.001\%^{+0.013\%}_{-0.001\%}$ for the physical experiments, logical experiments with error correction, and logical experiments with error detection, respectively, as illustrated in Fig. 2. This corresponds to a statistically significant reduction in the observed error rate by factors of 9.8 and 500 for the error corrected and the error detected experiments respectively.

B. Carbon code

We also entangled two pairs of logical qubits using a $[[12,2,4]]$ CSS code, which we nickname the *Carbon code* [41], and perform Bell correlation experiments with this entangled state. As a CSS code, Carbon has transversal CNOT and H gates, which immediately enable the preparation of a maximally entangled logical Bell state. However, the transversal CNOT gates create entanglement *between* qubit blocks, instead of entanglement *within* the block. For that reason, the Bell correlation experiment now involves two logical Bell states, as illustrated in Fig. 3.

1. Circuits

The Carbon code circuit for logical Bell state preparation requires 30 physical qubits (24 data qubits for the two blocks and 6 ancillas). While it is possible to fault-tolerantly prepare each block in tensor products of X or Z eigenstates and then prepare Bell states by applying
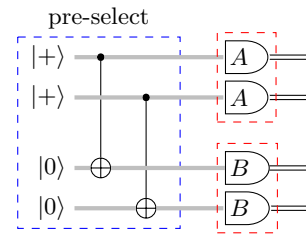

FIG. 3. Effective logical circuit for Bell state preparation using the $[[12,2,4]]$ Carbon code. The top and bottom pair of lines correspond to separate blocks. The pre-selected portion of the circuit (blue dashed box) include verification measurements at the physical level which are used for pre-selection (not shown). When post-selection is performed, it is based only on the syndrome information in the transversal measurements (red dashed boxes), and a separate decision is made for each code block. The logical observables A and B can be either X or Z independently, but for the experiments discussed here we focus on the scenario where $A = B$.

the transversal CNOT between them, it is more favorable to distribute the verification measurements throughout the Bell state preparation circuit (not shown). These verification circuits are tailored to the Bell state preparation, and can avoid measuring all stabilizer generators, focusing instead of simply detecting the propagation of failures into high weight Pauli errors. This simplification improves the fidelity of accepted state preparations in the experiments at the cost of mildly increasing the rejection rate.

We emphasize that the pre-selection criteria only looks to detect correlated error propagation in the unitary preparation circuits, as is customary in fault-tolerant stabilizer state preparation circuits. Most importantly, we do not measure the Bell correlations non-destructively in the preparation circuit, nor do we measure all stabilizer generators for the state.

The X and Z parities for each Bell pair can be obtained by measuring each block transversally, measuring both qubits within each block in the same basis [42]. As in Section III A, no error correction is performed between state preparation and measurement, but syndrome information from the transversal measurements allows us to detect errors nonetheless. Because the Carbon code has distance 4, while we can correct all weight 1 errors in each block, we cannot correct all weight 2 errors. However, we can *detect* all errors of weight up to 2, correct some of them with high probability, and reject the ones we cannot correct with high probability. Thus, the Carbon code experiments can have two modes of operation: an *error correction* mode, and an *error correction and rejection* mode. Importantly, error correction is performed in both modes of operation.

In error correction mode, we only perform pre-selection of state preparation, and do not allow any post-selection based on the syndrome information. Every experimental run that is accepted in pre-selection (or *pre-accepted*) yields logical experimental outcomes that may or may not

	runs	pre-accepted	post-accepted	corrections	errors	E_{xz} error rate	gain
unencoded baseline	$274,400$	—	—	—	$1,367$	$0.50\%^{+0.03\%}_{-0.03\%}$	—
encoded, pre-selection only	$24,200$	$18,035$	—	646	9	$0.05\%^{+0.04\%}_{-0.03\%}$	9.8
encoded, pre- and post-selection	$24,200$	$18,035$	$17,389$	—	0	$0.001\%^{+0.013\%}_{-0.001\%}$	500

TABLE I. Summary of experimental results for the preparation of Bell resource states using the $[[7,1,3]]$ Steane code and the corresponding circuit composed of physical qubits instead of logical qubits. Note this table only includes data for the X and Z basis measurements. Additional data including calculated process fidelities are given in Appendix C. We define "gain" to be the error rate of the physical circuits divided by the error rate of the corresponding logical circuit, while "corrections" refers to the number of pre-accepted experiments where a non-trivial syndrome was observed.

be corrected based on syndrome information. Decoding is performed by table look up after collecting the experimental data, and since we assume errors are independent and identically distributed, the table is populated sequentially by examining Pauli errors of increasing weight. If two Pauli errors of a particular weight have the same syndrome and their product is not in the stabilizer group, we choose arbitrarily which correction we associate with that syndrome, knowing that, for that particular syndrome we only make the right correction a constant fraction of the time. These resulting logical outcomes are then categorized as failures or successes based on the expected outcomes of a noiseless circuit (i.e., having parity outcome $+1$ for each Bell pair).

In the second mode of operation, we perform preselection of state preparation as before, but we also *post-select* on syndrome outcomes. Post-selection is based on the expected quality of decoding decisions. We modify the decoder table construction so that if two Pauli errors of a particular weight have the same syndrome and their product is not in the stabilizer group, the syndrome is added to the rejection set (so those outcomes are *post-rejected*), otherwise it is added to the acceptance set. For syndromes in the acceptance set, each syndrome is associated with the lowest weight Pauli error with that syndrome in the decoding look-up table [43].

The advantage of post-selecting on some syndrome outcomes for even distance codes is that it effectively boosts the distance of the code from d to $d+1$, in the sense that errors of weight $\frac{d}{2}$ do not cause logical errors anymore (as those are rejected in post-selection) [44]. Although postselection is not scalable in a strict sense, with low rejection rates one can run relatively deep circuits and achieve significant improvements to the logical error rate [44, 45]. The key observation is that the post-rejection rate is second order in the physical error rate, since weight 1 errors are corrected instead of rejected, and the constant factor is improved by correcting any unambiguous syndrome of an error with weight 2.

For the experiments here, the pre-rejection rates are a limiting factor, since the pre-rejection rate is first order in the physical error rate. However, with state preparation factories in a larger system, pre-selection is scalable, and post-selection becomes the dominant source of overhead.

2. Experimental results

We ran a total of 16,000 unencoded experiments, and 22,000 encoded experiments [46]. In both cases half of the runs measured X parities, and half of the runs measured Z parities.

For the unencoded experiments, out of the 16,000 runs, 125 yielded the incorrect parity, resulting is a physical error rate of $0.8\%^{+0.1\%}_{-0.1\%}$.

For the encoded experiments, the pre-selection procedure reduced the data set to a total of 15,483 out of the 22,000 encoded runs, roughly equally distributed across X and Z parities (a pre-acceptance rate of roughly 70%). Out of the 15,483 pre-accepted runs, 928 had a non-trivial syndrome, triggering a correction from the decoder, and in 26 experiments the resulting parity outcome was incorrect. When we allow for post-selection, the total number of accepted runs is reduced to 15,409, and 854 of these runs trigger a correction by the decoder (74 of the runs that triggered ambiguous syndromes were rejected by post-selection), and the logical parity was correct in all accepted runs. Note that the post-rejection rate is roughly 0.5%, two orders of magnitude lower than the pre-rejection rate, matching our expectations. These experimental results are summarized in Table II and Fig. 4.

Using the methods described in Appendix A, we estimate the error rate for the unencoded circuit is $0.8\%^{+0.1\%}_{-0.1\%}$, while our estimate for the encoded experiments with preselection is $0.17\%^{+0.07\%}_{-0.06\%}$ (a reduction in error rate by a factor of 4.7). Our estimate for the encoded experiments with pre-selection and post-selection is $0.001\%^{+0.015\%}_{-0.001\%}$, corresponding to a reduction in error rate by a factor of 800.

IV. REPEATED FAULT-TOLERANT ERROR CORRECTION

It is not sufficient to show shallow logical circuits that outperform their physical counterparts. The promise of quantum computers lies with solving large practical problems that require deep quantum circuits [8, 9]. This, in turn, requires fault-tolerant gadgets that intermix quantum error correction with fault-tolerant logical opera-

	runs	pre-accepted	post-accepted	corrections	errors	E_{xz} error rate	gain
unencoded baseline	16,000	16,000	—	—	125	$0.8^{+0.1}_{-0.1}\%$	—
encoded, pre-selection only	22,000	15,483	—	928	26	$0.17^{+0.07}_{-0.06}\%$	4.7
encoded, pre- and post-selection	22,000	15,483	15,409	854	0	$0.001\%^{+0.015\%}_{-0.001\%}$	800

TABLE II. Summary of experimental results from Bell correlation experiments with the $[[12, 2, 4]]$ Carbon code. We define "gain" to be the error rate of the unencoded circuits divided by the error rate of the encoded circuit in question, while "corrections" refers to how often a non-trivial syndrome was observed in the experiments that were accepted and pre- and/or post-selection.
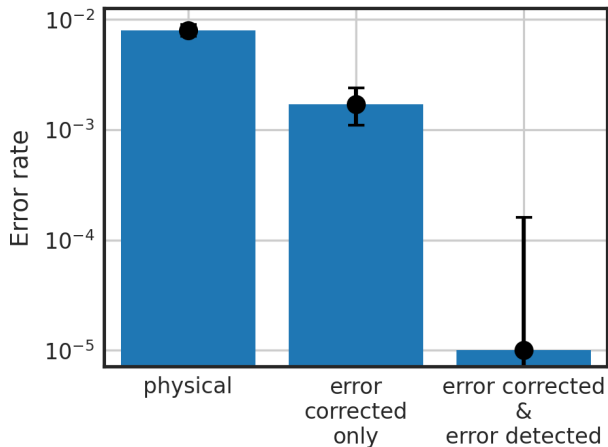

FIG. 4. Comparison between physical and logical error rates of the $[[12, 2, 4]]$ Carbon code for the Bell state preparation circuit in Fig. 3. Precise numerical values can be found in Table II. The difference in the error rates of the physical versus logical levels is statistically significant, as illustrated by the separation between the 95% confidence intervals.

tions [3, 5–7]. Composable, repeatable error correction is one of the key milestones in this journey, and only recently experimental demonstrations have become possible [13, 47, 48]. Two notable demonstrations from the last several years deserve special mention. A demonstration for the $[[7, 1, 3]]$ Steane code achieved $\approx 1.75\%$ logical error rate per error correction cycle for up to 6 rounds of error correction [12], a logical error rate just under one order of magnitude higher than the dominating error rate for the physical operations. Repeated error correction was also demonstrated for the surface code [13], achieving error rates of $\approx 9\%$ to $\approx 15\%$ per error correction cycle for up to 8 rounds of error correction, which is again just under one order of magnitude higher than the dominating error rate for the physical operations.

Here we demonstrate as many as 3 rounds of error correction for the $[[12, 2, 4]]$ Carbon code, using a combination of error correction and rejection. The Carbon code has a high threshold of approximately 1% and a rate of $\frac{1}{6}$ at distance 4 [49], which makes it a good candidate for experimental demonstrations. We compare the error rates of logical circuits with error correction and physical circuits, and show small circuits with logical error rates lower than the physical error rates by a statistically significant margin. We also give evidence that the error rate accumulated per round of error correction is comparable to the error rate accumulated with two physical CNOT in series.

A. Carbon code circuits

In the spirit of Ref. 22, instead of simply showing individual logical circuits that outperform their physical counterparts, we would like to show that families of circuits composed to perform some computation benefit from encoding. A crucial element of such a demonstration is repeated fault-tolerant error correction.

Despite the appealing features of the Steane code (such as transversal Clifford group and relatively compact state preparation and syndrome extraction circuits), estimates for the error threshold of the code remain relatively low, making experimental demonstrations of logical circuits outperforming physical circuits challenging.

The Carbon code, on the other hand, has a high threshold under the usual assumptions about parallelism, access to fresh ancillas, uniformity of the noise [49]. This proposal relies on teleportation-based syndrome extraction to achieve this performance [50]. Such an approach requires access to 3 code blocks within the hardware (one block for the data, and two blocks for the maximally entangled resource state).

Since the Carbon code requires 12 qubits per block, three blocks of the Carbon code would not fit in the 32-qubit H2 quantum processor [21]. Instead, we use a more compact scheme similar to Steane syndrome extraction [51], but derived from Knill's approach to syndrome extraction. The key observation made in Ref. 49 is illustrated in Fig. 5, resulting in a circuit that essentially serializes the syndrome extracting teleportation into two 1-bit teleportations [52]—the only requirement being that one must be able to prepare the encoded $|0\rangle$ and $|+\rangle$ states, apply an encoded CNOT, and measure in the encoded X and Z eigenbases (and all of these operations are available in the Carbon code).
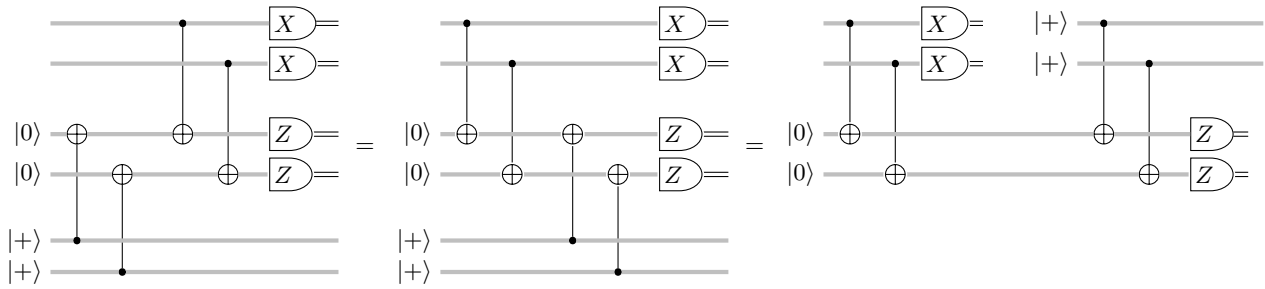
FIG. 5. Syndrome information can be obtained by performing teleportation at the logical level, as described by Knill [50]. Taking the realization of the original teleportation circuit for two logical qubits encoded in a Carbon block, which required 3 encoded blocks (left), it is possible to rearrange commuting circuit components to arrive at a circuit that uses a sequence of two 1-bit teleportations [52] to extract syndrome information requiring only 2 encoded blocks at any given time (right) [49].

B. Physical baselines

Aside from syndrome extraction, we would like to demonstrate encoded gates acting on the encoded information. Using the resources available in the experiment, we may only consider operations acting on a single data block, as the ancilla block and additional physical ancillas used in state preparation (31 qubits in total) take up most of the remaining qubit resources available in H2.

Within a single block, the simplest fault-tolerant operations available for the Carbon code are the $H \otimes H$ gates within a block, and a CNOT between the two logical qubits within the block—the H is implemented via transversal H application followed by a qubit permutation, while the CNOT is implemented via a qubit permutation [53, 54]. Since qubit permutations can be applied in the H2 system without interaction between qubits, the permutations do not require additional ancillas to be fault-tolerant.

Combining gate teleportation and permutation-based gates allows gates to perform logical circuits of the form depicted in Fig. 6 in a single round of error correction. The CNOTs and SWAPs of logical qubits within a block and their composition can be implemented by qubit permutations.

With the circuits of Figs. 5 and 6 in mind, we consider two physical baselines for the error correction experiments: a sequence of two physical 1-bit teleportations, and a sequence of two CNOTs (since we are focused on complete circuits [22], the inputs are fixed to be a tensor product of a pair of eigenstates of X or a pair of eigenstates of Z, and we choose the final measurement to be the corresponding tensor product observable).

The comparison between physical 1-bit teleportation and logical 1-bit teleportation informs us about the improvement error correction, pre-, and post-selection are providing over the additional encoding overhead, but without using the rigorous fault-tolerant gadget translation of the circuits from Ref. 7. In other words, this first baseline is a precursor for a comparison of fault-tolerant gadgets, and intuitively we expect that logical improvement over this baseline is necessary for an improvement with gadgets from Ref. 7.

The comparison between a sequence of two CNOTs and a round of error correction is motivated by the proposal in Ref. 22. Instead of benchmarking the entire circuit family implied by Fig. 6, we choose to compare against circuits two CNOTs since these are the physical gates with highest error rates in the physical system. Although in a fully scalable sub-threshold setting all logical gates can be made to improve upon their physical counterparts (by going to sufficiently high distance), the first bar that an experimental demonstration must meet is the demonstration of *some* non-trivial logical circuit that improves over its physical counterpart, which is our aim here.

C. Compiler optimizations

Time spent shuttling and cooling ions can lead to significant memory error, especially in error correction circuits that tend to leave some qubits idling for extended times [12]. Effective memory errors in H2 were previously reported in Ref. 21 for certain types of circuits with random connectivity, but the circuits used in this study are highly structured, leading to more coherent accumulation of noise, which is especially difficult to model accurately. To mitigate this problem, for the Carbon code experiments, we (1) modified the optimization of transport operations in the compiler and (2) incorporated dynamical decoupling pulses into the compiled circuits.

For the first optimization, we used a new cost function to find the optimal qubit assignments and transport routing, and we allowed the optimizer to run longer than it normally would during commercial operations. The new transport optimization cost function was created to better account for the time spent during gating relative to the time spent re-arranging qubits. The net result of this change was that more gates could be done in parallel reducing the number of gating steps at the expense of slightly less optimal re-arrangement. This optimization resulted in an approximate 15% reduction in the syndrome extraction time, significantly reducing memory errors. For the second optimization, the output of the compiler above, which contains the scheduling informa-
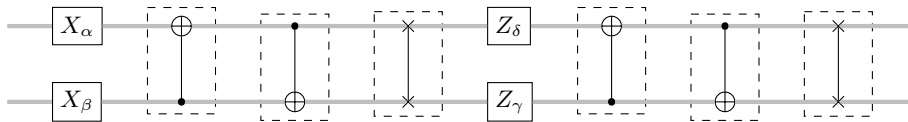
FIG. 6. Logical circuit family that can be implemented in a modified single round of error correction with the $[[12, 2, 4]]$ Carbon code, assuming $(\alpha, \beta, \delta, \gamma) \in \{0, \pm\frac{\pi}{2}, \pm\pi\}$. The logical single-qubit rotations require modified state preparation during syndrome extraction, and the order of X and Z rotations can be transposed by modifying the order of the 1-bit teleportations, but neither of these modification impact the ability to extract syndrome information. Any combination of gates in the dashed boxes can be implemented implemented by a permutation of the physical qubits.

tion for transport and quantum operations, was altered by inserting dynamical decoupling pulses into the schedule opportunistically to ensure no reschedule of operations or additional transport was needed. Both of these techniques are in experimental phases and not yet available to most H-series users. However, especially for the circuits presented in Section IV, the observed post-selection rejection rates for the complete circuits improved by $2\times$.

D. Experimental results

We observe a gain of more than an order of magnitude for a single round of error correction, as evidenced by a logical error rate of 0.03%, while the two physical baselines yield error rates of 0.51% (for physical teleportations) and 0.35% (a sequence of 2 CNOTs). For two rounds of error correction, the logical error rates increase to 0.4%, while the two physical baselines have error rates increase to 0.9% and 0.8% respectively, which represents a logical gain of 2 over the physical baselines. For three rounds of error correction the uncertainty in the logical error rate estimates do now allow us to conclusively point to an improvement (the error bars for the logical error rate estimate the and physical baseline error rates overlap slightly due to a limitation in the number of runs). Details for each experiment and associated statistical uncertainties can be found in Table III, and the results are illustrated in Fig. 7.

Although the gains over the physical error rates appear to behave non-linearly, the trend for the observed error rate per round of error correction is consistent with the expected linear behavior for a small number of repetitions, as illustrated in Fig. 7. The maximum-likelihood fits yield error rates per round of the teleportation and the CNOT baseline of $0.42\% \pm 0.09\%$ and $0.41\% \pm 0.08\%$ respectively, while the fit for each round of error correction points to $0.41\% \pm 0.13\%$—indicating that the differences in the error rates per round are not statistically significant.

The apparent non-linear trend in the gain is an artifact of the different y-intercepts for the linear trends of each of the sets of experiments. For the physical experiments, the y-intercept is set by the physical state-preparation and measurement (SPAM) error rate. This is also the case for the logical circuits, but details of the circuit implementation leads to a negative y-intercept despite positive SPAM error rates. This can be explained by

considering optimizations in scheduling of the operations in the experimental device. In particular, since memory and transport errors are the dominating source of error in the experiments (see Section II), we estimate the optimized compiled circuit to have roughly $2r - 1$ periods of significant waiting and transport for r rounds of error correction (due details of the initial state preparation and final state measurement), so the logical error rate for r rounds is estimated to be roughly

$$p_L(2r - 1) + p_{\text{SPAM},L} = 2\, p_L\, r + (p_{\text{SPAM},L} - p_L), \quad (1)$$

where p_L is the logical error rate per memory/transport period, and $p_{\text{SPAM},L}$ is the logical SPAM error rate. Since we expect that the logical error rate per memory/transport period is larger than the logical SPAM error rate, the intercept for the overall circuit logical error rate should be negative, agreeing with our experimental observations.

Surprisingly, there is no significant difference in the error rate per round of the two physical baselines, although absolute error rates are slightly more favorable for the physical CNOT baseline, likely due to it having fewer physical measurements.

The pre- and post-rejection rates also serve to inform whether the error correction circuits are behaving as expected. In particular we expect the pre- and post-rejection rates to be linear in the number of repetitions (to leading order), and the experimental data is highly consistent with that prediction (see Appendix B).

V. SUMMARY AND OUTLOOK

We have demonstrated several fault-tolerant circuits outperforming their physical counterparts in a state-of-the-art QCCD trapped-ion system. Using different codes and protocols, we demonstrated error rates 4.7 to 800 times lower than the physical error rates in Bell correlation experiments with the $[[7, 1, 3]]$ and the $[[12, 2, 4]]$ codes. Moreover, we demonstrated up to 3 rounds of error correction for the $[[12, 2, 4]]$ quantum code by combining error correction and error rejection, with logical error rates below a physical baseline circuit of two CNOTs per round. The error rate per round of error correction is consistent with error rates that are comparable with the physical error rates of two CNOTs, and potentially better,

		runs	pre-accepted	post-accepted	corrections	errors	error rate	gain
1 round	unencoded baseline (2 teleports)	20,000	—	—	—	102	$0.51\%^{+0.11\%}_{-0.09\%}$	—
	unencoded baseline (2 CNOTs)	20,000	—	—	—	70	$0.35\%^{+0.09\%}_{-0.08\%}$	—
	encoded, pre- and post-selection	10,000	7,054	7,022	1432	2	$0.03\%^{+0.06\%}_{-0.03\%}$	12–17
2 rounds	unencoded baseline (4 teleports)	20,000	—	—	—	189	$0.9\%^{+0.1\%}_{-0.1\%}$	—
	unencoded baseline (4 CNOTs)	20,000	—	—	—	154	$0.8\%^{+0.1\%}_{-0.1\%}$	—
	encoded, pre- and post-selection	14,548	7,685	7,345	2,948	28	$0.4\%^{+0.2\%}_{-0.1\%}$	2–2.25
3 rounds	unencoded baseline (6 teleports)	20,000	—	—	—	271	$1.4\%^{+0.2\%}_{-0.2\%}$	—
	unencoded baseline (6 CNOTs)	20,000	—	—	—	236	$1.2\%^{+0.2\%}_{-0.1\%}$	—
	encoded, pre- and post-selection	10,000	4,010	3,645	1,953	28	$0.8\%^{+0.3\%}_{-0.3\%}$	1.5–1.75

TABLE III. Summary of experimental results for 1 to 3 rounds of error correction via 1-bit teleportations using the $[[12, 2, 4]]$ Carbon code.
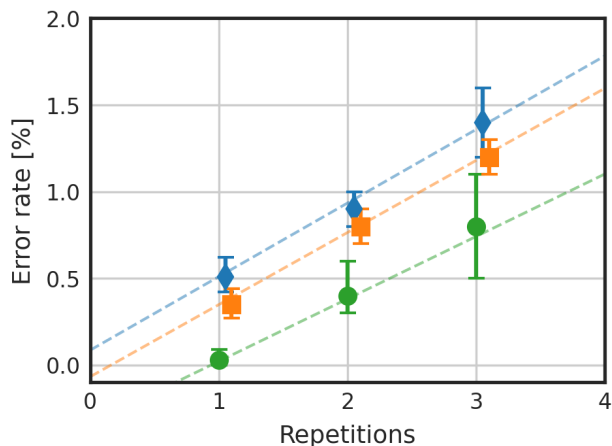

FIG. 7. Observed error rate for circuits with 1 to 3 rounds of error correction with the $[[12, 2, 4]]$ Carbon code (green circles) and physical baselines (blue diamond for pairs of 1-bit teleportations, and orange squares for pairs of CNOTs). Results are offset along the x-axis for clarity. Linear fits are obtained by maximum-likelihood estimation (see Appendix A for details).

although more data would be necessary to state this conclusively. With these results, we have demonstrated that current quantum processors are already able to reduce error rates in small circuits through quantum error correction. Future work will focus on extending these results to a richer set of fault-tolerant logical gadgets, and to ultimately enable universal fault-tolerant quantum computation while continuing to lower the achievable logical error rates. A significant milestone will be to demonstrate a universal family of quantum circuits with logical error rates approaching 10^{-8}.

ACKNOWLEDGMENTS

[1] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[2] A. M. Steane, Simple quantum error-correcting codes, Phys. Rev. A **54**, 4741 (1996).

[3] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '97 (Association for Computing Machinery, New York, NY, USA, 1997) p. 176–188.

[4] A. Y. Kitaev, Quantum error correction with imperfect gates, in *Quantum Communication, Computing, and Measurement* (Springer, 1997) pp. 181–188.

[5] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, Science **279**, 342 (1998).

[6] B. M. Terhal and G. Burkard, Fault-tolerant quantum computation for local non-markovian noise, Phys. Rev. A **71**, 012336 (2005).

[7] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, Quantum Info. Comput. **6**, 97–165 (2006), arXiv:quant-ph/0504218.

[8] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proceedings of the National Academy of Sciences **114**, 7555 (2017), arXiv:1605.03590 [quant-ph].

[9] M. E. Beverland, P. Murali, M. Troyer, K. M. Svore, T. Hoefler, V. Kliuchnikov, G. H. Low, M. Soeken, A. Sundaram, and A. Vaschillo, Assessing requirements to scale to practical quantum advantage, arXiv:2211.07629 [quant-ph] (2022).

[10] R. Raussendorf and J. Harrington, Fault-tolerant quantum computation with high threshold in two dimensions, Phys. Rev. Lett. **98**, 190504 (2007).

[11] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, Nature **598**, 281 (2021).

[12] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of real-time fault-tolerant quantum error correction, Phys. Rev. X **11**, 041058 (2021).

[13] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, A. D. T. Barba, S. Demura, A. Dunsworth, D. Eppens, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, L. F. Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, A. G. Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. P. Harrigan, S. D. Harrington, O. Higgott, J. Hilton, M. Hoffmann, S. Hong, T. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, K. Kechedzhi, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. Lill, W. Liu, A. Locharla, E. Lucero, F. D. Malone, J. Marshall, O. Martin, J. R. McClean, T. Mccourt, M. McEwen, A. Megrant, B. M. Costa, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, E. Mount, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, A. Opremcak, J. Platt, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, P. Roushan, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, V. Shvarts, J. Skruzny, V. Smelyanskiy, W. C. Smith, G. Sterling, D. Strain, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. V. Heidweiller, T. White, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, and N. Zhu, Suppressing quantum errors by scaling a surface code logical qubit, Nature **614**, 676 (2023), arXiv:2207.06431 [quant-ph].

[14] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. L. Brock, A. Z. Ding, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Real-time quantum error correction beyond breakeven, Nature **616**, 50–55 (2023), arXiv:2211.09116 [quant-ph].

[15] A. Erhard, H. P. Nautrup, M. Meth, L. Postler, R. Stricker, M. Ringbauer, P. Schindler, H. J. Briegel, R. Blatt, N. Friis, and T. Monz, Entangling logical qubits with lattice surgery, Nature **589**, 220 (2021), arXiv:2006.03071 [quant-ph].

[16] L. Postler, S. Heußen, I. Pogorelov, M. Rispler, T. Feldker, M. Meth, C. D. Marciniak, R. Stricker, M. Ringbauer, R. Blatt, P. Schindler, M. Müller, and T. Monz, Demonstration of fault-tolerant universal quantum gate operations, Nature **605**, 675 (2022), arXiv:2111.12654 [quant-ph].

[17] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, Logical quantum processor based on reconfigurable atom arrays, Nature **626**, 58–65 (2023), arXiv:2312.03982 [quant-ph].

[18] C. Ryan-Anderson, N. C. Brown, M. S. Allman, B. Arkin, G. Asa-Attuah, C. Baldwin, J. Berg, J. G. Bohnet, S. Braxton, N. Burdick, J. P. Campora, A. Chernoguzov, J. Esposito, B. Evans, D. Francois, J. P. Gaebler, T. M. Gatterman, J. Gerber, K. Gilmore, D. Gresh, A. Hall, A. Hankin, J. Hostetter, D. Lucchetti, K. Mayer, J. Myers, B. Neyenhuis, J. Santiago, J. Sedlacek, T. Skripka, A. Slattery, R. P. Stutz, J. Tait, R. Tobey, G. Vittorini, J. Walker, and D. Hayes, Implementing fault-tolerant entangling gates on the five-qubit code and the color code, arXiv:2208.01863 (2022).

[19] K. M. Svore, Defining logical qubits: Criteria for Resilient Quantum Computation (2023), (alt link) [Online; accessed 30-March-2024].

[20] J. Haah, What is Your Logical Qubit? (2024), (alt link) [Online; accessed 30-March-2024].

[21] S. A. Moses, C. H. Baldwin, M. S. Allman, R. Ancona, L. Ascarrunz, C. Barnes, J. Bartolotta, B. Bjork, P. Blanchard, M. Bohn, J. G. Bohnet, N. C. Brown, N. Q. Burdick, W. C. Burton, S. L. Campbell, J. P. Campora, C. Carron, J. Chambers, J. W. Chan, Y. H. Chen, A. Chernoguzov, E. Chertkov, J. Colina, J. P. Curtis, R. Daniel, M. DeCross, D. Deen, C. Delaney, J. M. Dreiling, C. T. Ertsgaard, J. Esposito, B. Estey, M. Fabrikant, C. Figgatt, C. Foltz, M. Foss-Feig, D. Francois, J. P. Gaebler, T. M. Gatterman, C. N. Gilbreth, J. Giles, E. Glynn, A. Hall, A. M. Hankin, A. Hansen, D. Hayes, B. Higashi, I. M. Hoffman, B. Horning, J. J. Hout, R. Jacobs, J. Johansen, L. Jones, J. Karcz, T. Klein, P. Lauria, P. Lee, D. Liefer, S. T. Lu, D. Lucchetti, C. Lytle, A. Malm, M. Matheny, B. Mathewson, K. Mayer, D. B. Miller, M. Mills, B. Neyenhuis, L. Nugent, S. Olson, J. Parks, G. N. Price, Z. Price, M. Pugh, A. Ransford, A. P. Reed, C. Roman, M. Rowe, C. Ryan-Anderson, S. Sanders, J. Sedlacek, P. Shevchuk, P. Siegfried, T. Skripka, B. Spaun, R. T. Sprenkle, R. P. Stutz, M. Swallows, R. I. Tobey, A. Tran, T. Tran, E. Vogt, C. Volin, J. Walker, A. M. Zolot, and J. M. Pino, A racetrack trapped-ion quantum processor, Phys. Rev. X **13**, 041052 (2023).

[22] D. Gottesman, Quantum fault-tolerance in small experiments, arXiv:1610.03507 [quant-ph] (2016).

[23] Complete quantum circuits are circuits that prepare qubits in a fixed state, perform a sequence of gates, and measure one or more qubits to yield classical output bits.

[24] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011), arXiv:1104.4695 [quant-ph].

[25] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011), arXiv:1104.3835 [quant-ph].

[26] D. J. Wineland, C. Monroe, W. M. Itano, D. Leibfried, B. E. King, and D. M. Meekhof, Experimental issues in coherent quantum-state manipulation of trapped atomic ions, Journal of Research of the National Institute of Standards and Technology **103**, 259 (1998), arXiv:quant-ph/9710025.

[27] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Demonstration of the trapped-ion quantum-ccd computer architecture, Nature 10.1038/s41586-021-03318-4 (2020), arXiv:2003.01293 [quant-ph].

[28] Quantinuum: Access to the H-Series Quantum Computer, https://www.quantinuum.com/hardware#access (2024), [Online; accessed 30-March-2024].

[29] Azure Quantum, https://quantum.microsoft.com (2024), [Online; accessed 30-March-2024].

[30] QIR Alliance, https://www.qir-alliance.org/ (2024), [Online; accessed 30-March-2024].

[31] IARPA, ELQ—Entangled Logical Qubits (2024), (alt link) [Online; accessed 29-March-2024].

[32] D. Nigg, M. Mueller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt, Quantum computations on a topologically encoded qubit, Science **345**, 302 (2014), arXiv:1403.5426 [quant-ph].

[33] J. Hilder, D. Pijn, O. Onishchenko, A. Stahl, M. Orth, B. Lekitsch, A. Rodriguez-Blanco, M. Müller, F. Schmidt-Kaler, and U. G. Poschinger, Fault-tolerant parity read-out on a shuttling-based trapped-ion quantum computer, Phys. Rev. X **12**, 011032 (2022).

[34] H. Goto, Minimizing resource overheads for fault-tolerant preparation of encoded states of the steane code, Scientific reports **6**, 1 (2016).

[35] R. Chao and B. W. Reichardt, Fault-tolerant quantum computation with few qubits, npj Quantum Information **4**, 1 (2018), arXiv:1705.05365 [quant-ph].

[36] J. Preskill, Reliable quantum computers, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**, 385 (1998).

[37] S. Bravyi and A. Kitaev, Universal quantum computation with ideal clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[38] D. Bacon, S. T. Flammia, A. W. Harrow, and J. Shi, Sparse quantum codes from quantum circuits, IEEE Transactions on Information Theory **63**, 2464 (2017), arXiv:1411.3334 [quant-ph].

[39] D. Gottesman, Opportunities and challenges in fault-tolerant quantum computation (2022), arXiv:2210.15844 [quant-ph].

[40] N. Delfosse and A. Paetznick, Spacetime codes of clifford circuits (2023), arXiv:2304.05943 [quant-ph].

[41] Details of the code and construction of logical operations will be made available in Ref. 49.

[42] The measurement of Y parities requires more complex circuitry (effectively applying the S gate to change bases), so we leave these more complex experiments for future work.

[43] Other rules based on information about the error models bias are certainly possible, but were not considered for these experiments.

[44] P. Prabhu and B. W. Reichardt, Distance-four quantum codes with combined postselection and error correction, arXiv:2112.03785 [quant-ph] (2021).

[45] E. H. Chen, T. J. Yoder, Y. Kim, N. Sundaresan, S. Srinivasan, M. Li, A. D. Córcoles, A. W. Cross, and M. Takita, Calibrated decoders for experimental quantum error correction, Physical Review Letters **128**, 10.1103/physrevlett.128.110504 (2022), arXiv:2110.04285 [quant-ph].

[46] We also ran experiments where we measured XZ and ZX cross-parities for each Bell pair, and confirmed that the distribution was close to uniform.

[47] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, Repeated quantum error detection in a surface code, Nature Physics **16**, 875–880 (2020), arXiv:1912.09410 [quant-ph].

[48] N. Sundaresan, T. J. Yoder, Y. Kim, M. Li, E. H. Chen, G. Harper, T. Thorbeck, A. W. Cross, A. D. Córcoles, and M. Takita, Demonstrating multi-round subsystem quantum error correction using matching and maximum likelihood decoders, Nature Communications **14**, 10.1038/s41467-023-38247-5 (2023), arXiv:2203.07205 [quant-ph].

[49] A. Paetznick *et al.* (2024), (to appear).

[50] E. Knill, Scalable quantum computation in the presence of large detected-error rates (2003), arXiv:quant-ph/0312190 [quant-ph].

[51] A. M. Steane, Active stabilization, quantum computation, and quantum state synthesis, Phys. Rev. Lett. **78**, 2252 (1997), arXiv:quant-ph/9611027.

[52] X. Zhou, D. W. Leung, and I. L. Chuang, Methodology for quantum logic gate construction, Phys. Rev. A **62**, 052316 (2000), arXiv:quant-ph/0002039.

[53] J. Leon, Computing automorphism groups of error-correcting codes, IEEE Transactions on Information Theory **28**, 496 (1982).

[54] M. Grassl and M. Roetteler, Leveraging automorphisms of quantum codes for fault-tolerant quantum computation, in *2013 IEEE International Symposium on Information Theory* (2013) pp. 534–538.

[55] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed. (University Science Books, 1996).

[56] J. A. Hanley and A. Lippman-hand, If nothing goes wrong, is everything all right?, Journal of the American Medical Association **249**, 1743 (1983).

[57] E. Eypasch, R. Lefering, C. K. Kum, and H. Troidl, Probability of adverse events that have not yet occurred: a statistical reminder, British Medical Journal **311**, 619 (1995).

[58] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Physics Letters A **303**, 249 (2002).

Appendix A: Statistical analysis and error bars

Error bars for failure rates are computed by assuming each experimental run of a particular circuit is an independent and identically distributed Bernoulli trial, so that the total count of failures given the number of trials is binomially distributed. Since the binomial distribution has the beta distribution as a conjugate prior distribution

of the binomial parameter $0 \leq p \leq 1$, we estimate the error bars by updating a beta prior with the outcome of the experiments.

More precisely, taking the prior distribution

$$p \sim \mathrm{Beta}(\alpha, \beta), \tag{A1}$$

the posterior distribution after running N experiments with F failures is

$$p \sim \mathrm{Beta}(\alpha + F, \beta + N - F), \tag{A2}$$

by applying Bayes' theorem to update our belief about the parameter p. We take $(\alpha, \beta) = (\frac{1}{2}, \frac{1}{2})$, known as *Jeffreys' prior*.

Throughout, the point estimates we report are the medians of the posteriors, while the error bars correspond to the 2.5% and the 97.5% of the posteriors, yielding a 95% credible interval. Following standard practice [55], we set significant digits for these estimate based on the most significant digit of the lower bound of the credible interval.

These estimates are conservative, in the sense that they are generally biased away from 0 and 1. Notably, even if no failures are detected, the median of the posterior will be non-zero, so that our point estimates for the error probability are never 0. The upper bound of the credible interval is marginally more conservative than the maximum risk 95% confidence "rule of 3" estimate used in medical and engineering fields [56, 57].

The linear fits depicted in Fig. 7, are estimated by maximum likelihood fitting of the linear model parameters. The likelihood is given by the joint probability density function of the observations, which consists of the product of the beta-distributed posteriors as outlined above. The uncertainty in the fit parameters is too large for conclusive statements about how the average accumulated error rates per round compare, although it is apparent that the gap between them is not large.

Appendix B: Rejection rates in the Carbon code

While we are particularly interested in the behavior of logical error rates for the fault-tolerant circuits, the behavior of the rejection rates for pre-selection and post-selection also informs us about the performance of the fault-tolerant circuits.

First, under the assumption of independent errors at each circuit location, pre-selection rejections are expected to be first-order in the physical error rates (recall pre-selection is not a barrier to scalability, since state preparation factories can be used to increase the probability of successful preparation exponentially close to 1). When error correction and error rejection are combined as we did in Sections III B and IV, on the other hand, cause the post-selection rejections rates to be second-order in the physical error rates, and therefore much smaller.
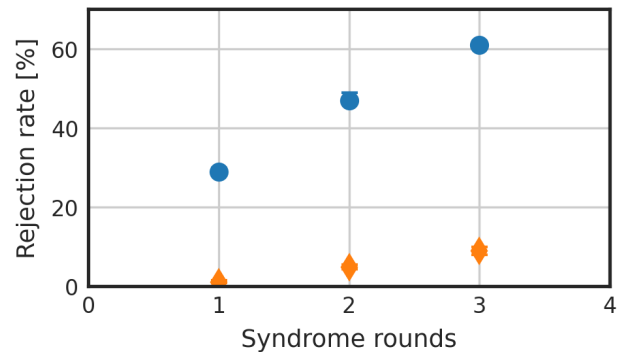

FIG. 8. Rejection rates during pre-selection (blue circles) and post-selection (orange diamonds) as a function of the number of error correction rounds for the $[[12, 2, 4]]$ Carbon code. In both cases the expected linear trend (as a function of number of repetitions) is apparent, as is the large separation between pre-rejections and post-rejections.

Second, for small circuits, the rejection rates should increase roughly linearly with the logical volume of the circuit. In our case, the logical volume is simply proportional to the number of error correction rounds.

These expected trends are apparent in Fig. 8, as is the large separation between the pre- and post-rejection rates, consistent with what would be expected between first- and second-order events.

Appendix C: Additional experimental data for the Steane code experiments

The Steane code admits transversal $Y \otimes Y$ measurements, allowing a straight-forward measurement of the process fidelity for the Bell state preparation. For a Bell pair, the average fidelity is given by $F_a = (I \otimes I + X \otimes X - Y \otimes Y + Z \otimes Z)/4$, and the process fidelity is $F_p = ((d+1)F_a - 1)/d$ [58], where the dimension $d = 4$ for the Bell pair. In this section, Table IV, Table V, and Table VI provide the process fidelities for the various Steane code experiments and their physical level analogs, and a break-down of the results into the three different Pauli measurements required.

Note that the "gain" values in Table V and Table VI are calculated by dividing the process fidelity (which is calculated using the $X \otimes X$, $Y \otimes Y$, and $Z \otimes Z$ bases measurements) of the unencoded circuit by the process fidelity of the encoded circuit.

Pauli	runs	error rate
$X \otimes X$	$137,200$	$0.42\%^{+0.04\%}_{-0.03\%}$
$Y \otimes Y$	$137,200$	$0.39\%^{+0.03\%}_{-0.03\%}$
$Z \otimes Z$	$137,200$	$0.58\%^{+0.04\%}_{-0.04\%}$
E_{xz}	$274,400$	$0.50\%^{+0.03\%}_{-0.03\%}$
E_p	NA	$0.69\%^{+0.03\%}_{-0.03\%}$

TABLE IV. The measured process fidelity E_p of the physical-level Bell preparation gadget and the error rates of individual Pauli operators $X \otimes X$, $Y \otimes Y$, and $Z \otimes Z$. For comparison, the error rate determined from measuring the probability of getting the wrong parity from $X \otimes X$ and $Z \otimes Z$ is given as E_{xz}.

Pauli	runs	pre-accepted	error rate	gain
$X \otimes X$	$12,100$	$9,025$	$0.10\%^{+0.08\%}_{-0.05\%}$	4.2
$Y \otimes Y$	$12,100$	$9,082$	$0.09\%^{+0.08\%}_{-0.05\%}$	4.3
$Z \otimes Z$	$12,100$	$9,010$	$0.003\%^{0.025\%}_{-0.003\%}$	193
E_{xz}	$24,200$	$18,035$	$0.05\%^{+0.04\%}_{-0.03\%}$	9.8
E_p	NA	NA	$0.10\%^{+0.06\%}_{-0.04\%}$	6.8

TABLE V. The measured process fidelity E_p of the logical-level Bell preparation gadget utilizing the Steane code where the destructive measurements are analyzed using quantum error correction. The error rates of individual Pauli operators $X \otimes X$, $Y \otimes Y$, and $Z \otimes Z$ are also reported. For comparison, the error rate determined from measuring the probability of getting the wrong parity from $X \otimes X$ and $Z \otimes Z$ is given as E_{xz}.

Pauli	runs	pre-accepted	post-accepted	error rate	gain
$X \otimes X$	$12,100$	$9,025$	$8,688$	$0.003\%^{+0.026\%}_{-0.003\%}$	140
$Y \otimes Y$	$12,100$	$9,082$	$8,665$	$0.03\%^{+0.05\%}_{-0.02\%}\%$	13
$Z \otimes Z$	$12,100$	$9,010$	$8,701$	$0.003\%^{+0.026\%}_{-0.003\%}$	193
E_{xz}	$24,200$	$18,035$	$17,389$	$0.001\%^{+0.013\%}_{-0.001\%}$	500
E_p	NA	NA	NA	$0.02\%^{+0.03\%}_{-0.01\%}$	35

TABLE VI. The measured process fidelity E_p of the logical-level Bell preparation gadget utilizing the Steane code where the destructive measurements are analyzed using QED. The error rates of individual Pauli operators $X \otimes X$, $Y \otimes Y$, and $Z \otimes Z$ are also reported. For comparison, the error rate determined from measuring the probability of getting the wrong parity from $X \otimes X$ and $Z \otimes Z$ is given as E_{xz}.