


# Flag Fault-Tolerant Error Correction for any Stabilizer Code

Rui Chao<sup>\*</sup> and Ben W. Reichardt

*Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*

 (Received 24 January 2020; accepted 29 June 2020; published 3 September 2020)

Conventional fault-tolerant quantum error-correction schemes require a number of extra qubits that grow linearly with the code’s maximum stabilizer generator weight. For some common distance-three codes, the recent “flag paradigm” uses just two extra qubits. Chamberland and Beverland [Quantum 2, 53 (2018)] provide a framework for flag error correction of arbitrary-distance codes. However, their construction requires conditions that only some code families are known to satisfy. We give a flag error-correction scheme that works for any stabilizer code, unconditionally. With fast qubit measurement and reset, it uses  $\leq d + 1$  extra qubits for a distance- $d$  code.

DOI: [10.1103/PRXQuantum.1.010302](https://doi.org/10.1103/PRXQuantum.1.010302)

## I. INTRODUCTION

In quantum error correction, errors are diagnosed by measuring the code’s check operators. The measurement circuits themselves are faulty. One way to measure the syndromes fault tolerantly is to prepare some extra qubits in a special state, then couple these ancillas to the data qubits. Conventional fault-tolerant error-correction schemes [1–4] need as many ancillas as the maximum stabilizer generator weight. The “flag method” [5,6], however, requires only two ancilla qubits for common distance-three codes. Chamberland and Beverland [7] have generalized the flag idea to arbitrary-distance codes. Here, we give an explicit, ancilla-efficient flag fault-tolerant error-correction scheme that unconditionally applies to arbitrary stabilizer codes.

Figure 1 compares the flag method to the Shor [1] and DiVincenzo-Aliferis decoding [4] methods. In the flag method, the syndrome is extracted into a single ancilla, on which one gate failure can spread to two or more errors on the data. To fix this problem, the syndrome ancilla is protected with a flag gadget. The flag measurement signals the possible occurrence of any correlated errors. Provided that the flag circuit is carefully designed, the possible errors are distinguishable by their syndromes and so can be corrected.

Flag error-correction schemes have been given for several code families: Hamming codes [5], rotated surface and Reed-Muller codes [7], color codes [7–11], cyclic CSS

codes [12,13], and heavy hexagon and square codes [14]. Flag-based schemes have been proposed also for error detection [15], logical state preparation [16–18] and lattice surgery [19,20]. In particular, Chamberland and Beverland [7] have formalized the flag idea for error correction and extended it to arbitrary-distance codes. For a distance- $d$  code, they have shown how to construct a flag error-correction protocol, provided that the “detectability” and “distinguishability” requirements are satisfied. (Detectability requires that each stabilizer generator is measured with a flag circuit that signals whenever  $s < d/2$  faults result in more than  $s$  data errors. Distinguishability requires that any two data error patterns, each arising from  $< d/2$  faults spread among the stabilizer generators’ flag circuits and resulting in the same flag pattern, either have distinct syndromes or differ by a stabilizer.)

Here, we provide a flag error-correction scheme that applies to any stabilizer code, unconditionally. In particular, the circuit construction relies only on the code’s distance and stabilizer generator weights. The key idea is that harmful faults can be approximately located from the observed flag pattern and then corrected right away, before they propagate to correlated data errors. For a distance- $d$  code with stabilizer generators of maximum weight  $w$ , the scheme requires  $\min\{d + 1, 2w + 2\}$  extra qubits. No further qubits are needed, provided that the ancillas can be reset rapidly, in time comparable to a gate operation.

Different fault-tolerant error-correction schemes are summarized in Table I. Unlike Shor’s and similar schemes [1–3], our procedure does not verify the ancillas; no postselection is needed. Unlike previous flag schemes [5,7,12], ours is nonadaptive, in the sense that the flag-measurement outcomes alone provide enough information to correct correlated errors; no extra syndrome measurement is required. Prabhu and Reichardt [22] have recently

<sup>\*</sup>ruichao@usc.edu

*Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.*

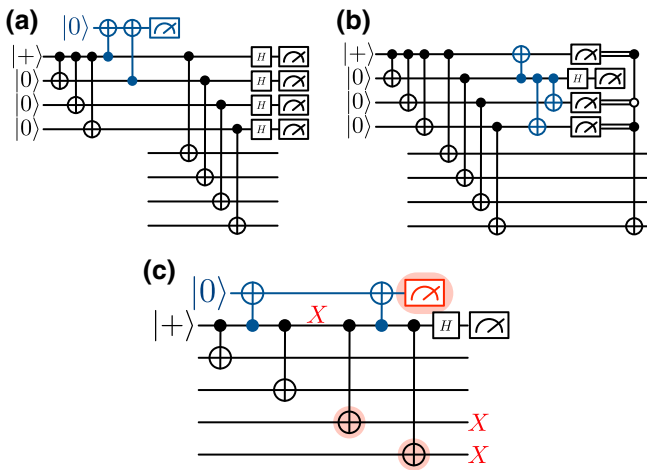


FIG. 1. Different fault-tolerant methods for measuring  $X^{\otimes 4}$ . (a) Shor’s method. Prepare a four-qubit cat state and then detect correlated errors using an extra ancilla qubit (blue). Apply subsequent transversal gates to the data only when the detection outcome is trivial. (b) DiVincenzo-Aliferis decoding method. Couple the data qubits with an unverified cat state. Then apply the decoding circuit (blue) followed by qubitwise measurements, the outcomes of which give the syndrome and information about potential correlated data errors. This method has been extended to measure an arbitrary-weight stabilizer of a distance-three code (Supplemental Material of Ref. [5]). (c) Flag method. Extract the syndrome into an ancilla qubit, which is protected by a flag gadget (blue). A single fault, e.g.,  $X$ , on the syndrome ancilla can propagate into correlated data errors, e.g.,  $X^{\otimes 2}$ . Such correlated errors are signaled by a nontrivial measurement outcome on the flag ancilla, and can be diagnosed using subsequent syndrome measurements.

provided new flag error-correction schemes that apply to arbitrary distance-three and -five codes, and require, respectively,  $O(\log w)$  and  $O(w)$  ancilla qubits, where  $w$  is the maximum stabilizer generator weight. Their schemes are also nonadaptive but allow slow qubit reset; each ancilla qubit is measured only once per stabilizer generator.

## II. FAULT-TOLERANCE DEFINITIONS

**Definition 1** ([23,24]). *An error-correction protocol tolerates  $t$  faults if it satisfies the following.*

- (1) If the input has  $r$  errors and the protocol has  $s$  faults with  $r + s \leq t$ , then the output is correctable by perfect decoding.
- (2) If the protocol has  $s$  faults with  $s \leq t$ , then the output is at most  $s$  errors away from some codeword.

Assume that faults can occur in elementary operations: single-qubit preparations and measurements, single- and two-qubit gates, and idle positions.

TABLE I. Numbers of ancilla qubits for different error-correction schemes. Here,  $d$  is the code distance and  $w$  is the maximum stabilizer generator weight. In Refs. [5], [7], and [12], flag patterns can only detect or reveal partial information about the correlated data errors; adaptive syndrome measurement is further needed for full correction. In Refs. [1–3], correlated data errors cannot occur due to ancilla verification. In the other schemes in the table, syndrome-measurement circuits can be scheduled deterministically; (flag) ancilla patterns provide enough information to correct correlated errors.

	Adaptive	Nonadaptive
Slow qubit reset	None	$\lceil w/2 \rceil$ for $d = 3$ [4,5,21] $O(\log w)$ for $d = 3$ [22] $\lceil w/2 \rceil + O(1)$ for $d = 5$ [22]
Fast qubit reset	$O(1)$ for many codes [5,7,12]	$\Omega(w)$ [1–3] $d + 1$ for all $d$ (this paper)

Our protocol measures one stabilizer generator at a time. The *flag circuit* uses one *syndrome ancilla* and one or more *flag ancillas*. If there are no faults, measuring the syndrome ancilla gives the syndrome, and the flag ancilla measurement outcomes will be trivial.

**Definition 2.** *A flag syndrome-measurement circuit is nonadaptive and  $t$ -fault-tolerant ( $t$ -FT) if any  $k \leq t$  faults result in at most  $k$  data errors, after corrections based on the flag pattern.*

An important observation is that one can construct an error-correction protocol tolerating  $t$  faults by measuring one stabilizer generator at a time, each using a  $t$ -FT flag circuit. Indeed,  $t$ -FT circuits automatically guarantee condition 1 of Definition 1. One way to achieve condition 2 is to keep measuring the syndromes until the same syndrome repeats consecutively a certain number of times without flags being triggered [1].

Notice that Definition 2 is stronger than that of a  $t$ -flag circuit, introduced by Chamberland and Beverland [7], which guarantees a nontrivial flag pattern whenever  $s \leq t$  faults result in more than  $s$  errors. (In particular, Ref. [7] gives a scheme for constructing a 2-flag circuit that requires five ancilla qubits; and gives a conjectured scheme for constructing a general  $t$ -flag circuit that requires  $w$  ancilla qubits.)

Denote by  $|f|$  the Hamming weight, and by  $f \oplus g$  the bitwise XOR, of binary strings  $f$  and  $g$ . Single-qubit measurements in diagrams are all in the  $Z$  basis ( $|0\rangle, |1\rangle$ ).

## III. SYNDROME MEASUREMENT FOR ANY STABILIZER CODE

We now present a flag circuit that measures a stabilizer of an arbitrary distance- $d$  code. Without loss of generality, we may take the stabilizer to be  $X^{\otimes w}$ . Since

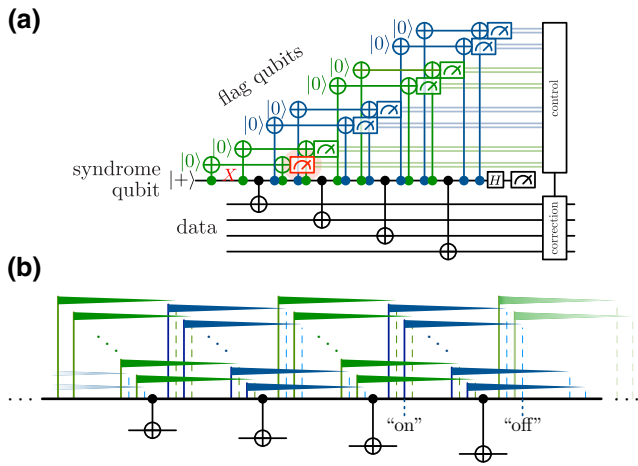


FIG. 2. (a) Distance-three fault-tolerant measurement of  $X^{\otimes 4}$ . Each controlled-NOT (CNOT) gate to the data is protected by two flags.  $X$  corrections are applied based on the flag pattern. (b) Schematic showing how the circuit extends to  $X^{\otimes w}$  and arbitrary distance  $d$ ; control and corrections are omitted. Between each consecutive pair of CNOT gates to the data,  $2t$  flags are turned on, and  $2t$  turned off, in an alternating fashion: on, off, on, off.

measuring  $X^{\otimes w}$  can incur at most  $w$  data errors, let  $t = \min \{\lfloor d - 1/2 \rfloor, w\}$ .

Figure 2 illustrates the construction. A syndrome ancilla, initialized in  $|+\rangle$ , is coupled to the  $w$  data qubits with CNOT gates. Each of these CNOT gates is protected by a different set of  $2t$  flags. Between CNOT gates to the data, the flags are carefully transitioned—with flag  $j$  for the next CNOT turned on just before flag  $j$  for the previous CNOT is turned off, for  $j \in \{1, 2, \dots, 2t\}$ —so that there are always either  $2t$  or  $2t + 1$  active flags, except at the very beginning and end. Observe that with fast qubit reset, this construction requires at most  $2t + 1$  flag ancillas, so  $2t + 2$  ancilla qubits total.

**Conjecture 3.** *With appropriate correction rules based on the flag pattern, the procedure of Fig. 2 is  $t$ -FT.*

A stronger statement might also hold. Based on computer analyses of circuits for  $t \in \{1, 2, 3\}$  (see Appendix), it appears that the data CNOTs toward the beginning or end require fewer flags:  $0, 1, 2, \dots, 2t - 1, 2t, 2t, \dots, 2t, 2t - 1, \dots, 2, 1, 0$  flags instead of always  $2t$  flags. The first and last data CNOTs require no flag protection because a syndrome-ancilla fault immediately before or after these gates can propagate to an error of weight only 0 or 1. Moving toward the middle, the number of flags needed then steadily increases up to a maximum of  $2t$ . Figure 3 shows an example, with  $w = 10$  and  $t = 2$ .

However, we do not attempt to prove Conjecture 3. Instead, in order to simplify the correction rules' boundary conditions, modify the scheme of Fig. 2 by prepending  $r = t(t + 2)$  rounds of  $2t$  flags. That is, follow Fig. 2 for

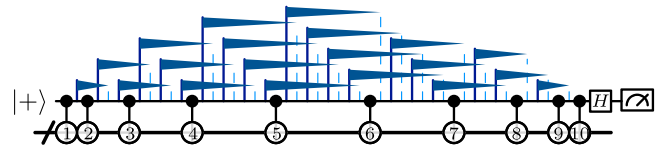


FIG. 3. Optimized ( $t = 2$ )-FT circuit for measuring  $X^{\otimes 10}$ . Note that the central data CNOT gates are protected by  $2t$  flags, but fewer flags are used toward the ends. With appropriate correction rules, the circuit tolerates up to two faults.

a weight  $r + w$  stabilizer, except do not use data CNOTs in the first  $r$  rounds. We show the following.

**Lemma 4.** *The modified procedure, with  $t(t + 2)$  initial flag rounds and appropriate correction rules, is  $t$ -FT.*

Let us first give some intuition. Distinguish possible circuit faults into syndrome faults and flag faults. Syndrome faults cause  $X$  faults on the syndrome ancilla; they can spread to correlated data errors. Flag faults are measurement faults on flag ancillas; they do not spread to the data. A syndrome fault can trigger  $2t$  or  $2t + 1$  flags, as in Fig. 4(a), or  $<2t$  flags on the left and right boundaries, as in Fig. 2(a). The motivation for the initial flag rounds is so that our correction algorithm, which works from left to right, does not need to account for left boundary conditions.

Assume for the moment that every syndrome fault flips exactly  $2t$  flags. Imagine laying the flags along the real

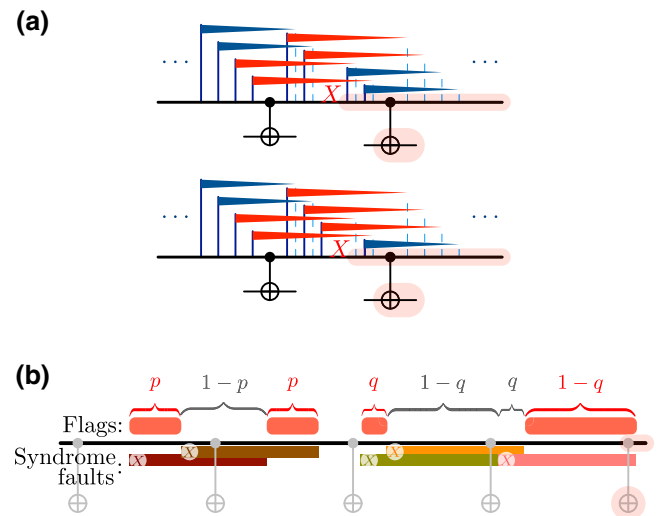


FIG. 4. (a) A syndrome fault can flip  $2t$  or  $2t + 1$  flags. (b) Syndrome faults spread to the data and flip intervals of flags. The correction algorithm uses the observed flag pattern, which can be further distorted by flag faults, to infer the syndrome faults' approximate locations.

line, so that successive flag measurements cover consecutive intervals of length  $1/2t$ ; then each syndrome fault flips an interval of length one. The up to  $t$  flag faults can further flip intervals with total length up to  $t/2t = 1/2$ .

As shown in Fig. 4(b), the pattern of triggered flags depends on how the flipped intervals overlap. The correction algorithm needs to use the observed flags to deduce the approximate locations of the syndrome faults. It does not need to determine the exact locations because all that matters is how the faults propagate to data-qubit errors, via the downward data CNOTs located at integers on the line. Also, an even number of syndrome faults with overlapping intervals, as on the left side of Fig. 4(b), does not require any data correction for fault tolerance.

**Proof of Lemma 4.** We give an explicit correction algorithm and then analyze it inductively.

Since a syndrome fault can trigger flags in up to two consecutive flag rounds, with at least  $t(t-1+2) + t = t(t+2)$  initial flag rounds, there must be at least one sequence of at least  $t$  consecutive flag rounds in which no flags are triggered. Call this sequence the “start” rounds. Having identified the start rounds, the algorithm can discard all flags to their left. Indeed, faults to the left of the start rounds cause no data errors, and do not affect the flag pattern past the start rounds. [As on the right side of Fig. 4(b), three syndrome faults cancel out over a length-one interval; more generally  $t$  faults can cancel out an interval of length at most  $t/2 < t$ . Faults cannot cancel out over such a long stretch as the  $t$  start rounds.]

For fault tolerance, it suffices to consider only syndrome and flag faults; any other fault either can be ignored or is equivalent to a syndrome or flag fault. Divide the faults after the start rounds into three sets: set  $F$  of flag faults, set  $S_{2t+1}$  of syndrome faults flipping exactly  $2t+1$  flags, and set  $S_{2t}$  of syndrome faults either flipping  $2t$  flags or on the last round. They satisfy  $|F| + |S_{2t+1}| + |S_{2t}| \leq t$ . For instance, a  $YY$  fault on a data CNOT is in  $S_{2t}$ .

Now we specify our correction algorithm, and then prove that syndrome measurement is  $t$ -FT.

The algorithm will work through the flags from the start rounds on, from left to right, interpreting one round of  $2t$  flags at a time. Denote by  $f_1, f_2, \dots \in \{0, 1\}^{2t}$  the rounds of flags after the start rounds,  $f_\ell$  being the flags around data (or dummy) CNOT  $\ell$ , as in Fig. 5. Let  $\Omega$  be the rightmost nontrivial round, and define  $\varepsilon \in \{0, 1\}^{2t}, |\varepsilon| = 1$  as the rightmost nontrivial flag of  $f_\Omega$ .

---

```

 $m_0 \leftarrow 0^{2t}$ 
for  $i = 1, 2, \dots, \Omega$  do
   $m_i \leftarrow m_{i-1} \oplus f_i$ 
  if  $|m_i| > t$  or  $(i = \Omega, |m_i| = t, |m_i \oplus \varepsilon| > t)$  then
     $m_i \leftarrow m_i \oplus 1^{2t}$  corner-case fix
    Apply  $X$  to data CNOT targets from  $i$  onwards

```

---

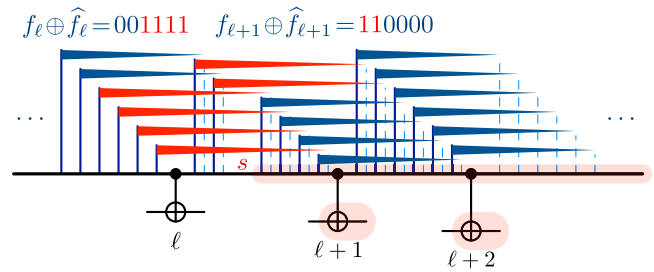


FIG. 5. If  $t = 3, s \in S$  flips six flags in  $f_\ell$  and  $f_{\ell+1}$ .

The algorithm deals with each round in the same way except round  $\Omega$ . When  $i = \Omega$  and  $|m_i| = t$ , the corner-case fix effectively flips the rightmost nontrivial flag  $\varepsilon$ . (Roughly,  $m_i$  stores the running total, mod 2, of the flag rounds, up to the corrections that flip everything.)

The proof proceeds by incorporating faults one at a time but with  $\Omega$  and  $\varepsilon$  fixed. We compare the algorithm’s behavior before and after each fault is incorporated. For fault tolerance, we argue that the residual data errors after the corrections never outnumber the current faults.

Split each fault in  $S_{2t+1}$  into a  $2t$  bit syndrome fault, plus one extra flag fault at the end. That is, define  $\bar{F}$  to be the union of  $F$  and the rightmost bits from  $S_{2t+1}$ , and let  $S$  be the union of  $S_{2t}$  and remaining parts of  $S_{2t+1}$ . (Thus,  $|\bar{F}| = |F| + |S_{2t+1}|, |S| = |S_{2t}| + |S_{2t+1}|$ .)

For the base case, we consider all the faults in  $\bar{F}$  with no data errors. Define  $r_1, \dots, r_\Omega \in \{0, 1\}^{2t}$  as the flag patterns flipped by faults in  $\bar{F}$ . In subsequent inductive steps, we add faults in  $S$ , from left to right, one at a time. Suppose the added fault  $s$  occurs between data CNOTs  $\ell$  and  $\ell + 1$ , causing data errors  $P_\ell X_{\ell+1} X_{\ell+2} \dots$  for some  $P \in \{I, X, Y, Z\}$ . We prove, by induction in  $s$ , that for all  $i \geq \ell + 1$ , the value of  $m_i$  when the algorithm finishes satisfies  $m_i = \bigoplus_{j=1}^i r_j$ . (For the base case, set  $\ell = 0$ .)

Consider the base case of faults in  $\bar{F}$ . Then  $f_i = r_i$ . We claim that  $m_i = \bigoplus_{j=1}^i r_j$ . This is clear since  $|\bigoplus_{j=1}^i r_j| \leq |\bar{F}| \leq t$  and so corrections are never triggered except possibly at round  $\Omega$ . If  $|m_{\Omega-1} \oplus r_\Omega| = t$ , then necessarily  $S_{2t} = \emptyset$  and the flag  $\varepsilon$  is attributed to the rightmost fault in  $\bar{F}$ . In particular,  $|m_\Omega \oplus \varepsilon| = t - 1$ . Therefore,  $|m_i|$  or  $|m_\Omega \oplus \varepsilon|$  never exceeds  $t$ , and the algorithm applies no corrections, as desired.

Now consider adding a syndrome fault  $s \in S$ , between data CNOTs  $\ell$  and  $\ell + 1$ . Define  $\hat{f}_i$  and  $\hat{m}_i$  as the parameters of the algorithm (when finished) in the previous induction step, without  $s$ . (See Fig. 5 for an example.) With  $s$  added, the algorithm’s behavior does not change until  $i = \ell$ . (Recall that  $\Omega$  and  $\varepsilon$  are fixed.) Beyond  $\ell$ , the following observation essentially finishes the proof.

**Observation.** In exactly one of the two rounds  $i = \ell$  and  $\ell + 1$  the algorithm changes its behavior: either applying



the correction when it did not before, or vice versa. The algorithm's behavior remains unchanged for  $i \geq \ell + 2$ .

Since  $s$  flips  $2t$  flags, we have

$$(\widehat{f}_\ell \oplus f_\ell) \oplus (\widehat{f}_{\ell+1} \oplus f_{\ell+1}) = 1^{2t}. \quad (1)$$

Note that  $m_\ell \oplus m_{\ell-1} \oplus f_\ell \in \{1^{2t}, 0^{2t}\}$ , depending on whether in round  $\ell$ ,  $X$  data corrections are applied or not. To show the observation, we consider all possible scenarios.

(1)  $\widehat{m}_\ell \oplus \widehat{m}_{\ell-1} \oplus \widehat{f}_\ell \neq m_\ell \oplus m_{\ell-1} \oplus f_\ell$ , i.e., the correction behavior changes at round  $\ell$ . Since  $\widehat{m}_{\ell-1} = m_{\ell-1}$ , we have  $\widehat{m}_\ell \oplus \widehat{f}_\ell \oplus m_\ell \oplus f_\ell = 1^{2t}$ . Adding Eq. (1),  $m_\ell \oplus f_{\ell+1} = \widehat{m}_\ell \oplus \widehat{f}_{\ell+1}$ , and the algorithm's behavior does not change from round  $\ell + 1$  on. In particular, for  $i \geq \ell + 1$ , still  $m_i = \widehat{m}_i = \bigoplus_{j=1}^i r_j$ . The change of corrections  $X_\ell X_{\ell+1} \dots$  leaves at most one data error  $P_\ell X_\ell$ , as desired.

(2)  $\widehat{m}_\ell \oplus \widehat{m}_{\ell-1} \oplus \widehat{f}_\ell = m_\ell \oplus m_{\ell-1} \oplus f_\ell$ , i.e., the correction behavior does not change at round  $\ell$ . Adding  $\widehat{m}_{\ell-1} = m_{\ell-1}$  and Eq. (1) gives  $(\widehat{m}_\ell \oplus \widehat{f}_{\ell+1}) \oplus (m_\ell \oplus f_{\ell+1}) = 1^{2t}$ . We claim the correction behavior changes at round  $\ell + 1$ , and so the change of corrections  $X_{\ell+1} X_{\ell+2} \dots$  leaves at most one data error  $P_\ell$ , as desired.

(a)  $|\widehat{m}_\ell \oplus \widehat{f}_{\ell+1}| \neq t$ . Then one of the terms  $\widehat{m}_\ell \oplus \widehat{f}_{\ell+1}$  and  $m_\ell \oplus f_{\ell+1}$  has weight  $> t$  and the other  $< t$ . The correction behavior changes in round  $\ell + 1$ . Thus, we have  $\widehat{m}_{\ell+1} \oplus m_{\ell+1} = (\widehat{m}_\ell \oplus \widehat{f}_{\ell+1}) \oplus (m_\ell \oplus f_{\ell+1}) \oplus 1^{2t} = 0^{2t}$  and so  $m_i = \widehat{m}_i = \bigoplus_{j=1}^i r_j$  for  $i \geq \ell + 1$ .

(b)  $|\widehat{m}_\ell \oplus \widehat{f}_{\ell+1}| = t$ . Then  $|\widehat{m}_{\ell+1}| = t$ . Then since  $\widehat{m}_{\ell+1} = \bigoplus_{j=1}^{\ell+1} r_j$  and  $|\bigoplus_{j=1}^{\ell+1} r_j| \leq |\overline{F}|$ , we have  $S_{2t} = \emptyset$ . Therefore,  $\ell + 1 = \Omega$  and the corner-case fix is triggered in round  $\ell + 1$ , both with and without  $s$ . Since  $(\widehat{m}_\ell \oplus \widehat{f}_{\ell+1}) \oplus (m_\ell \oplus f_{\ell+1}) = 1^{2t}$  and  $|\varepsilon| = 1$ ,  $\{|\varepsilon \oplus m_\ell \oplus f_{\ell+1}|, |\varepsilon \oplus \widehat{m}_\ell \oplus \widehat{f}_{\ell+1}|\} = \{t-1, t+1\}$ . As one of the terms has weight  $> t$ , the correction behavior changes in round  $\ell + 1$ .

When  $s$  occurs after the last data CNOT, it can flip fewer than  $2t$  flags—but since it causes no data errors, it can be ignored. ■

Note that at the end  $|m_\Omega|$  lower bounds the number of faults. This information can be used to reduce the measurement repetitions in error correction [1,7,25].

#### IV. CONCLUSION

Intuitively, our protocol might be optimal in the number of ancilla qubits, but it is open to prove this. Our scheme is most practical for quantum devices with good memory and connectivity, such as ion traps [26]. Good memory allows ancillas to be reset without introducing excessive noise on waiting qubits, and good connectivity allows the data and

flag qubits to be coupled to the syndrome qubit. Following Fig. 3, our scheme can be adapted to specific codes. More work is needed to adapt it to stricter geometry constraints, as in superconducting qubits [27].

#### ACKNOWLEDGMENTS

Research supported by NSF Grant No. CCF-1254119, ARO Grant No. W911NF-12-1-0541, MURI Grant No. FA9550-18-1-0161 and a Google Quantum award.

#### APPENDIX: COMPUTER SEARCH FOR CORRECTION RULES

Given an arbitrary flag circuit that measures a stabilizer generator of a specific code, it is generally difficult to tell whether there exist correction rules to achieve fault tolerance. The pseudocode below runs a brute-force search for  $t$ -FT correction rules, given a flag circuit  $C$  as input. Without loss of generality, the stabilizer is taken to be  $X^{\otimes w}$ .

In the pseudocode,  $\text{flag}(F)$  and  $\text{error}(F)$  denote the flag pattern and data error pattern resulting from a set  $F$  of circuit faults, respectively. In particular, the  $i$ th bit of  $\text{error}(F) \in \{0, 1\}^w$  is 1 if and only if the  $i$ th data qubit is afflicted by a nontrivial Pauli.

Recall that  $C$  consists of only one syndrome ancilla, possibly multiple flag ancillas, and only CNOT gates. Let  $\mathcal{F}$  be a set of faults in  $C$  and consist of  $X$  faults just before every data or flag CNOT control as well as  $X$  faults on every flag measurement. Then, any relevant fault in  $C$  is equivalent to some element of  $\mathcal{F}$ , except for two-qubit faults on data CNOTs, with  $X$  on the control and  $P \in \{I, X, Y, Z\}$  on the target. These two-qubit faults are accounted for with a subset  $\mathcal{F}^* \subseteq \mathcal{F}$ , which consists of  $X$  faults just before every data CNOT control.

$\mathcal{D}$  is a dictionary whose keys  $k$  are possible flag patterns  $\text{flag}(F)$  generated by combinations  $F$  of at most  $t$  faults from  $\mathcal{F}$ . The value  $v$  for each key consists of all the combinations of at most  $t$  faults that generate the corresponding flag pattern. Then, there exist  $t$ -FT correction rules for  $C$  if and only if for each key-value pair  $(k, v) \in \mathcal{D}$  there exist data corrections consistent for all the fault combinations in  $v$ .

---

Input:  $t, C$

Let:  $\mathcal{F}$  = set of faults in  $C$  consisting of

- $X$  faults just before every data or flag CNOT control
- $X$  faults on every flag measurement

Let:  $\mathcal{F}^* \subseteq \mathcal{F}$  be the faults before the data CNOT controls

$\mathcal{D} \leftarrow \{(k, v) \mid v = \{F \subseteq \mathcal{F} \mid \text{flag}(F) = k, |F| \leq t\}\}$

$\mathcal{R} \leftarrow \emptyset$

for  $(k, v)$  in  $\mathcal{D}$

  for  $F$  in  $v$

$\mathcal{Q} \leftarrow \{\text{data qubits not targeted by CNOTs in } F \cap \mathcal{F}^*\}$

$\mathcal{S}_F \leftarrow \{\text{error}(F) \oplus E \mid E \in \{0, 1\}^w, |E_{\mathcal{Q}}| \leq |\mathcal{F} \setminus \mathcal{F}^*|\}$

$\mathcal{R}[k] \leftarrow \bigcap_{F \in v} \mathcal{S}_F$

  if  $\mathcal{R}[k] = \emptyset$

    return Fail

---

Specifically, for each  $F \in \mathcal{V}$  of a given pair  $(k, \mathcal{V})$ , denote by  $\mathcal{S}_F$  the set that contains all the valid  $X$  data corrections for  $F$ , in form of binary strings. To calculate  $\mathcal{S}_F$ , consider the subset  $\mathcal{Q} \subseteq \{1, 2, \dots, w\}$  that contains all the data qubits not targeted by any CNOT from  $F \cap \mathcal{F}^*$ . Observe that data errors on  $\mathcal{Q}$  are solely  $X$  errors, and the complement  $\overline{\mathcal{Q}} = \{1, 2, \dots, w\} \setminus \mathcal{Q}$  contains all the data qubits afflicted by unknown Paulis. Further denote by  $E \in \{0, 1\}^w$  the residual data errors of  $F$  after possible  $X$  data corrections  $\text{error}(F) \oplus E$ . In particular, the  $i$ th bit of  $E$  is 1 if and only if the  $i$ th data qubit is left with a nontrivial Pauli after a possible  $X$  correction indicated by the  $i$ th bit of  $\text{error}(F) \oplus E$ . Then, in the worst case the residual data errors  $E_{\overline{\mathcal{Q}}}$  restricted to  $\overline{\mathcal{Q}}$  have weight  $|F \cap \mathcal{F}^*|$ . Hence for fault tolerance, corrections on  $\mathcal{Q}$  should remove sufficient  $X$  errors so that the residual data errors  $E_{\mathcal{Q}}$  restricted to  $\mathcal{Q}$  have weight no greater than  $|F \setminus \mathcal{F}^*|$ .

If for any  $(k, \mathcal{V})$  the intersection  $\bigcap_{F \in \mathcal{V}} \mathcal{S}_F$  is empty, then the search fails and  $C$  is not  $t$ -FT. Otherwise, the code returns a dictionary  $\mathcal{R}$  of valid correction rules indexed by the observed flag pattern  $k$ .

- 
- [1] P. W. Shor, in *Proceedings of the 37th Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, NY, 1996), p. 96.
- [2] A. M. Steane, Active Stabilization, Quantum Computation, and Quantum State Synthesis, *Phys. Rev. Lett.* **78**, 2252 (1997).
- [3] E. Knill, Scalable quantum computing in the presence of large detected-error rates, *Phys. Rev. A* **71**, 042322 (2005).
- [4] D. P. DiVincenzo and P. Aliferis, Effective Fault-Tolerant Quantum Computation with Slow Measurements, *Phys. Rev. Lett.* **98**, 020501 (2007).
- [5] R. Chao and B. W. Reichardt, Quantum Error Correction with Only Two Extra Qubits, *Phys. Rev. Lett.* **121**, 050502 (2018).
- [6] R. Chao and B. W. Reichardt, Fault-tolerant quantum computation with few qubits, *npj Quant. Inf.* **4**, 42 (2018).
- [7] C. Chamberland and M. E. Beverland, Flag fault-tolerant error correction with arbitrary distance codes, *Quantum* **2**, 53 (2018).
- [8] B. W. Reichardt, Fault-tolerant quantum error correction for Steane's seven-qubit color code with few or no extra qubits, [arXiv:1804.06995](https://arxiv.org/abs/1804.06995) (2018).
- [9] P. Baireuther, M. D. Caio, B. Criger, C. W. J. Beenakker, and T. E. O'Brien, Neural network decoder for topological color codes with circuit level noise, *New J. Phys.* **21**, 013003 (2019).
- [10] A. Bermudez, X. Xu, M. Gutiérrez, S. C. Benjamin, and M. Müller, Fault-tolerant protection of near-term trapped-ion topological qubits under realistic noise sources, *Phys. Rev. A* **100**, 062307 (2019).
- [11] C. Chamberland, A. Kubica, T. Yoder, and G. Zhu, Triangular color codes on trivalent graphs with flag qubits, *New J. Phys.* **22**, 023019 (2020).
- [12] T. Tansuwannont, C. Chamberland, and D. Leung, Flag fault-tolerant error correction, measurement, and quantum computation for cyclic Calderbank-Shor-Steane codes, *Phys. Rev. A* **101**, 012342 (2020).
- [13] T. Tansuwannont and D. Leung, Fault-tolerant quantum error correction using error weight parities, [arXiv:2006.03068](https://arxiv.org/abs/2006.03068) (2020).
- [14] C. Chamberland, G. Zhu, T. J. Yoder, J. B. Hertzberg, and A. W. Cross, Topological and Subsystem Codes on Low-Degree Graphs with Flag Qubits, *Phys. Rev. X* **10**, 011022 (2020).
- [15] C. Vuillot, Is error detection helpful on IBM 5Q chips? *Quant. Inf. Comput.* **18**, 949 (2018).
- [16] C. Chamberland and A. W. Cross, Fault-tolerant magic state preparation with flag qubits, *Quantum* **3**, 143 (2019).
- [17] Y. Shi, C. Chamberland, and A. Cross, Fault-tolerant preparation of approximate GKP states, *New J. Phys.* **21**, 093007 (2019).
- [18] C. Chamberland and K. Noh, Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits, [arXiv:2003.03049](https://arxiv.org/abs/2003.03049) (2020).
- [19] M. Gutiérrez, M. Müller, and A. Bermudez, Transversality and lattice surgery: Exploring realistic routes toward coupled logical qubits with trapped-ion quantum processors, *Phys. Rev. A* **99**, 022330 (2019).
- [20] L. Lao and C. G. Almudever, Fault-tolerant quantum error correction on near-term quantum processors using flag and bridge qubits, *Phys. Rev. A* **101**, 032333 (2020).
- [21] A. M. Stephens, Efficient fault-tolerant decoding of topological color codes, [arXiv:1402.3037](https://arxiv.org/abs/1402.3037) (2014).
- [22] P. Prabhu and B. W. Reichardt, Fault-tolerant syndrome measurement with fewer qubits (to be published).
- [23] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, *Quant. Inf. Comput.* **6**, 97 (2006).
- [24] D. Gottesman, in *Quantum Information Science and its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics* (American Mathematical Society, Providence, RI, 2010), Vol. 68, p. 13.
- [25] C. Zalka, Threshold estimate for fault tolerant quantum computation, [axXiv:quant-ph/9612028](https://arxiv.org/abs/quant-ph/9612028) (1997).
- [26] C. Monroe and J. Kim, Scaling the ion trap quantum processor, *Science* **339**, 1164 (2013).
- [27] M. H. Devoret and R. J. Schoelkopf, Superconducting circuits for quantum information: an outlook, *Science* **339**, 1169 (2013).