Minimizing readout-induced noise for early fault-tolerant quantum computers

Yunzhe Zheng [1,2,*] and Keita Kanno [1]

[1]*QunaSys Inc., Aqua Hakusan Building 9F, 1-13-7 Hakusan, Bunkyo, Tokyo 113-0001, Japan*
[2]*QuTech, Delft University of Technology, Delft 2628 CJ, Netherlands*

Quantum error correcting code can diagnose potential errors and correct them based on measured outcomes by leveraging syndrome measurement. However, mid-circuit measurement has been technically challenging for early fault-tolerant quantum computers and the readout-induced noise acts as a main contributor to the logical infidelity. We present a different method for syndrome extraction, namely *generalized syndrome measurement*, that requires only a single-shot measurement on a single ancilla, while the canonical syndrome measurement requires multiple measurements to extract the eigenvalue for each stabilizer generator. As such, we can detect the error in the logical state with minimized readout-induced noise. By adopting our method as a precheck routine for quantum error correcting cycles, we can significantly reduce the readout overhead, the idling time, and the logical error rate during syndrome measurement. We numerically analyze the performance of our protocol using Iceberg code and Steane code under realistic noise parameters based on superconducting hardware and demonstrate the advantage of our protocol in the near-term scenario. As mid-circuit measurements are still error-prone for near-term quantum hardware, our method could boost the applications of early fault-tolerant quantum computing.

I. INTRODUCTION

Recent experimental progress has demonstrated powerful and scalable quantum computers with different hardware like superconducting qubits [1–7], neutral atoms [8–14], and trapped ions [15–17]. To release the full quantum power, practical quantum hardware needs to overcome various kinds of noise and neutralize their effect on the quantum states storing information. Quantum error correcting (QEC) codes provide a general framework to tackle noise by encoding a logical qubit into multiple physical qubits [18,19]. The extra Hilbert space allows the detection and correction of quantum errors, providing the possibility for large-scale quantum computing.

To diagnose physical errors in QEC codes, syndrome measurement (SM) is required to gain the necessary information for identifying both the location and type of errors. Let us consider a stabilizer code Q, which has a k-element stabilizer generator set $\{S_i\}$. For each S_i, canonical syndrome measurement requires preparation of an ancilla state in $|+\rangle$, followed by applying a Pauli S_i gate controlled by the ancilla and final measurement of the ancilla in the X basis [Fig. 1(a)]. However, as mid-circuit readouts have been extremely error prone for near-term quantum hardware, readout-induced noise significantly affects the fidelity of the logical state after the syndrome measurement. For superconducting qubits,

the state-of-the-art readout duration ranges from hundreds of nanoseconds to several microseconds while the coherence time for scalable near-term hardware is tens of microseconds [4,5]. When the ancilla qubit is being measured out, all data qubits are idling and therefore suffer from decoherence noise. Besides, the state-of-the-art readout assignment error is on the scale of 0.01, which cannot be mitigated if no repeated readouts are performed. For neutral atom arrays, on-site mid-circuit readouts are practically challenging and the ancilla qubit to be measured must be moved to a separate zone by coherent transportation to avoid unintended interaction on other physical qubits [8,13,14], which adds time overhead and might be susceptible to unexpected errors.

Exploiting redundant ancilla qubits and measuring different stabilizers in parallel may alleviate the readout-induced decoherence, but it poses a stricter requirement for qubit control and connectivity and may bring further space overhead for near-term quantum hardware. The direct readout error, e.g. assignment error and crosstalk, is still present even if every stabilizer is measured simultaneously with separate ancilla qubits. Therefore, we consider it favorable to reduce the number of readouts for better performance on early fault-tolerant quantum computers. Notably, there have been studies regarding measurement-free QEC protocols that could achieve fault-tolerant quantum computing without ancilla readout [20–23]. However, these protocols require the ability to reset the qubits, which is generally as noisy as readout in the near-term quantum hardware [13,24,25]. Therefore, it in principle just transfers the overhead from readout to resetting and only works for limited quantum hardware where resetting is much less error-prone than readout.

In this paper, we propose a different method to extract syndrome information, namely *generalized syndrome*
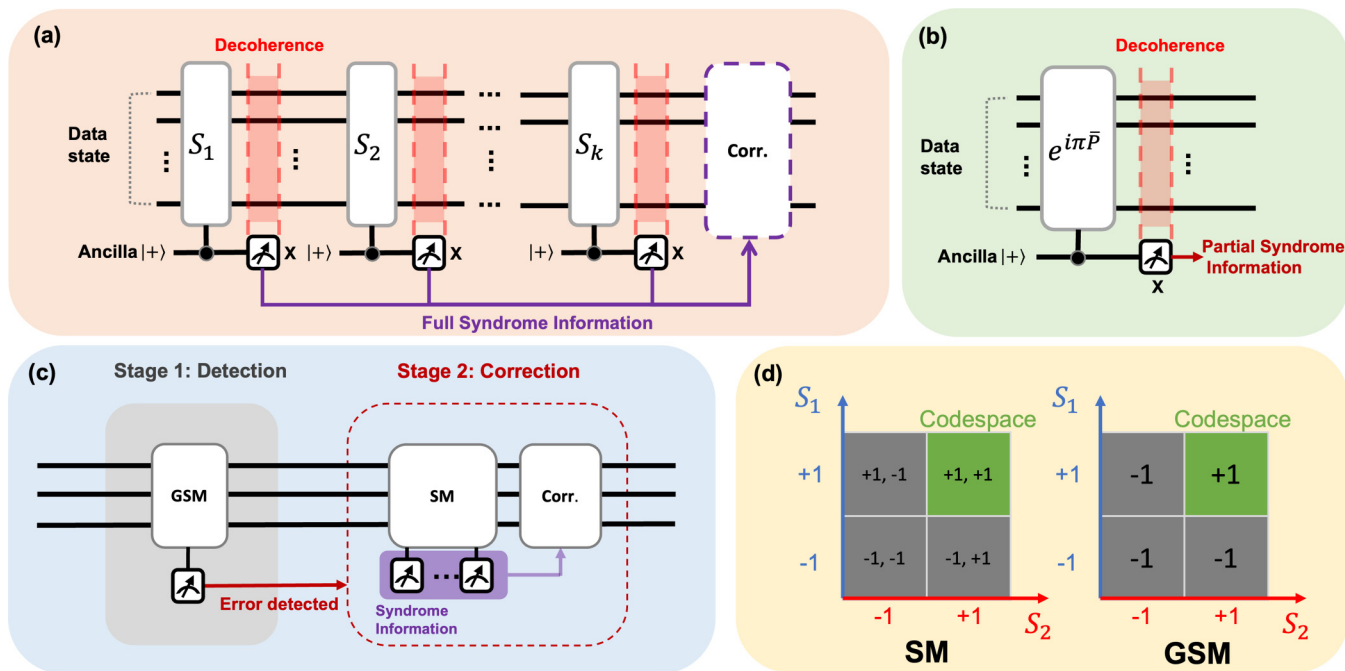
---

*dran.z@foxmail.com

FIG. 1. (a) Circuit scheme for canonical syndrome measurement (SM). A total number of k (the number of stabilizer generators in the code) readouts are required to diagnose the error syndrome and apply proper correcting operations. While the ancilla qubit is being measured, all data qubits are subjected to decoherence due to idling. (b) Circuit scheme for (one-shot) generalized syndrome measurement (GSM). Only a single readout is needed to extract the partial information and detect the error, and the readout-induced idling error is minimized. (c) QEC cycle with generalized syndrome measurement. We divide the cycle into two stages: detection and correction. We first use the GSM method to check whether an error has occurred or not. If no error is detected, we skip the correction stage while ensuring that the logical state is in the codespace. If an error is detected, we follow the GSM routine by the canonical SM, extracting the full syndrome information and correcting the detected error. (d) A two-stabilizer code example illustrates the difference between the SM method and the GSM method. The value within the box denotes the measured eigenvalue for S_i (SM) or the codespace projector $\bar{P}$ (GSM).

measurement (GSM), that requires only single measurement on a single ancilla regardless of the number of stabilizer generators. As such, our protocol would minimize the time of readout for syndrome check. Specifically, we check the encoded states by directly measuring the eigenvalue of the codespace projector instead of the eigenvalue of each stabilizer generator, which allows us to check whether an error occurred or not at a single-shot readout overhead. Further, we propose to exploit the GSM method as a precheck routine for canonical syndrome measurement in QEC cycles, which may render the following canonical routine unnecessary depending on the GSM readout outcome. As a result, our method can reduce the average number of readouts for quantum error correction, given that the input logical state only suffers from low level of noise. We numerically demonstrate the performance of our protocol with Iceberg code [26] and Steane code [27], and show that our method outperforms the canonical method in the practical noisy scenario. As long as mid-circuit readouts are still a main infidelity source for quantum hardware, we believe our method can reduce the impact of readout-induced noise and enhance the performance of quantum applications in early fault-tolerant quantum computing.

II. PROTOCOL DESCRIPTION

We first introduce the idea of the GSM method. For a QEC code Q with stabilizer generator $\{S_i\}(i = 1, 2, ..., k)$, each

stabilizer generator S_i has an associate projector $P_i = (I + S_i)/2$. We can define the codespace projector of Q,

$$\bar{P} = \prod_{i=1}^{k} P_i = \frac{1}{2^k} \prod_{i=1}^{k} (I + S_i). \tag{1}$$

Instead of measuring each stabilizer individually, we aim to detect whether a given state is in the codespace or not by measuring the eigenvalue of $\bar{P}$ using controlled unitary gates $e^{i\pi\bar{P}}$. As shown in Fig. 1(b), we start with the noisy input logical state $|\psi\rangle$ and an ancilla initiated at $|+\rangle_a$,

$$\frac{1}{\sqrt{2}}(|0\rangle_a|\psi\rangle + |1\rangle_a|\psi\rangle). \tag{2}$$

Next, we apply a control-$e^{i\pi\bar{P}}$ gate triggered by the ancilla state $|1\rangle_a$, and the premeasurement state becomes

$$\frac{1}{\sqrt{2}}(|0\rangle_a|\psi\rangle + |1\rangle_a e^{i\pi\bar{P}}|\psi\rangle). \tag{3}$$

Rewriting the ancilla state in the X basis, we have

$$(|+\rangle_a(I - \bar{P})|\psi\rangle + |-\rangle_a\bar{P}|\psi\rangle), \tag{4}$$

where we used the fact that $\bar{P}^2 = \bar{P}$ and $e^{i\beta\bar{P}} = I + (e^{i\beta} - 1)\bar{P}$ for any angle β. If we measure the ancilla qubit to be $|-\rangle_a$, the logical state will be projected to the state proportional to $\bar{P}|\psi\rangle$, which lies in the codespace of Q, and we assure there is no detectable error within the ability of code Q. Otherwise, the

postmeasurement state lies outside of the codespace, implying that at least one detectable error has occurred in the logical state.

Compared with the GSM method, the canonical SM method projects the input state to the codespace (+1 eigenspace of all S_i) by measuring the ancilla qubits sequentially. As each stabilizer S_i requires an ancilla readout, at least k readouts are required to tell whether the state is in the codespace or not and the readout-induced idling time will be kt_m given t_m is the time for a physical readout. The direct readout errors, like assignment errors and crosstalk errors, may also affect the fidelity of the logical state by k times. In the GSM method, we require only a single measurement to extract partial syndrome information and diagnose if an error exists, at the cost of losing the specific information of which kind of error exists. Therefore, the GSM method can take the place of the SM method for quantum error detection (QED), where we postselect logical states based on readout outcomes. Figure 1(d) provides an illustrative example of a code containing two stabilizer generators. For the purpose of QED, the GSM method cannot completely replace the canonical SM method as all syndrome information is necessary. However, the GSM method can still bring benefits for QEC cycles: In a fault-tolerant quantum computing regime, The logical error rate of the input state is generally very low to ensure that we are under the noise threshold where logical qubits would outperform physical qubits. Therefore, the main purpose of a QEC cycle is to assure that no error has occurred in the input logical state, rather than revealing the specific information about the presented errors. Based on this assumption, we can integrate the GSM method in QEC cycles in an adaptive way and divide a cycle into two stages as presented in Fig. 1(c): In the first stage, we detect if there is an error in the logical state using the GSM method. If no error is detected (in a more probable case), we assure the logical state is in the codespace, skip the next stage, and finish this QEC cycle. If at least an error is indeed detected (in a less probable case), we follow up with the canonical SM method to extract the full syndrome information and correct all detected errors. As long as the input state has sufficiently high fidelity, we can lower the average time of ancilla readouts by introducing the GSM method in such a combined way. For example, if the input state (encoded in a k-stabilizer code) is the ideal state with 99% probability or a fully erroneous state with 1% probability, the average readout-induced idling time for a noiseless QEC cycle with the GSM method can be approximated by $[0.99 \cdot 1 + 0.01 \cdot (1 + k)]t_m = (1 + 0.01k)t_m$, which is much shorter than the time kt_m for the canonical method even if $k = 2$.

Notably, the idea of adaptive syndrome measurement has already been used for Shor-style error correction [28,29], but we are using it here in a different way: The first measurement is for the whole codespace projector, and the following measurements are for each stabilizer generator. Our method could also be further made in Shor-style fault-tolerant way (see Appendix F). As mid-circuit readouts are still very error prone for current large-scale quantum hardware, we expect our method could significantly reduce the noise originating from readouts and benefit early fault-tolerant quantum applications.

In addition to the one-shot GSM method shown in Fig. 1(b) where we just use one ancilla readout, one can also use a few more ancilla readouts (but still less than the canonical SM method) for a more reduced gate complexity and relaxed connectivity requirement. This trade-off can be achieved by splitting the codespace projector $\bar{P}$ into several subprojectors $\{\bar{P}_1, \bar{P}_2, ..., \bar{P}_m\}(m < k)$ such that

$$\bar{P} = \prod_{i=1}^{m} \bar{P}_i \tag{5}$$

and

$$\bar{P}_i = \prod_{\{j\}_i} P_j, \tag{6}$$

where $\{j\}_i$ is an index set associated with each subprojector and $\bigcup_{i=1}^{m}\{j\}_i = \{x \in \mathbb{Z} | 0 \leqslant x \leqslant k\}$. By measuring these subprojectors $\bar{P}_i$ separately, we can achieve an m-shot GSM protocol where the total readout time is still less than the canonical SM method. For example, a four-stabilizer code (e.g. the five-qubit [[5, 1, 3]] code) with S_1, S_2, S_3, S_4 can exploit a two-shot GSM by splitting the general projector into two subprojectors, one containing S_1 and S_2 while the other containing S_3 and S_4. We give a more detailed explanation in the Appendix B.

III. GATE IMPLEMENTATION OF $e^{i\pi\bar{P}}$

In the GSM method, we demand the implementation for the controlled projector gates $e^{i\pi\bar{P}}$, which is a highly nonlocal unitary and requires to be decomposed into elementary gates to be implemented in near-term quantum hardware. Here, we provide two different methods of implementation.

In the first method of decomposition, we expand the expression of $\bar{P}$ in Eq. (1) into a sum of stabilizer operators M_i,

$$\bar{P} = \frac{1}{2^k} \prod_{i=1}^{k} (I + S_i) = \frac{1}{K} \sum_{i=1}^{K} M_i, \tag{7}$$

where $K = 2^k$ for simplicity. As S_i commutes with each other, M_i is also commutable with each other. Thus, an $e^{i\pi\bar{P}}$ gate can be written as

$$e^{i\pi\bar{P}} = \prod_{i=1}^{K} e^{i\pi M_i/K} = \prod_{i=1}^{K} \mathbf{M_i}(\pi/K), \tag{8}$$

where $\mathbf{M_i}(\theta) = e^{i\theta M_i}$ are multi-qubit Pauli phase gates with angle θ. The controlled projector gates are therefore rearranged in the gate sequences as shown in Fig. 2(b). As the identity gate I is also among $\mathbf{M_i}$, there are $K - 1$ nontrivial gates that need to be implemented in practice. This kind of implementation is particularly suitable for systems with globally tunable Hamiltonian, like trapped ions [31] or neutral atoms [32], but can also be decomposed into local Clifford and single-qubit phase gates [33]. Notably, although the number of multi-qubit Pauli gates grows exponentially with k, the total physical Hamiltonian evolution time is bounded as the rotation angle for each $\mathbf{M_i}$ also decays exponentially with k.
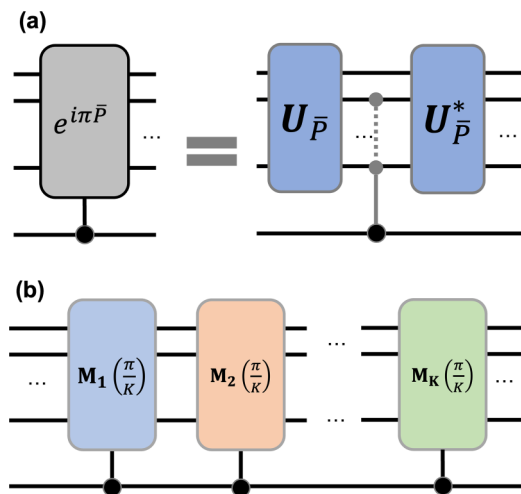
(a)

(b)

FIG. 2. Gate Implementation of $e^{i\pi\bar{P}}$. (a) Sandwich implementation. $U_{\bar{P}}$ is the circuit decoding the code with projector $\bar{P}$ back to the trivial code state, and there is a multi-qubit control phase gate sandwiched in the middle between $U_{\bar{P}}$ and its conjugate $U_{\bar{P}}^*$. (b) The $e^{i\pi\bar{P}}$ can also be implemented by a sequence of control multi-qubit Pauli rotation gates.

In the second method, we can construct the controlled $e^{i\pi\bar{P}}$ using a multi-qubit control phase gate sandwiched by a Clifford unitary $U_{\bar{P}}$ and its conjugate as shown in Fig. 2(a). The unitary $U_{\bar{P}}$ always exists as any $[[n, k, d]]$ stabilizer codes can be encoded by a Clifford unitary from the trivial code state

$$|\psi\rangle_{\text{data}} \otimes |1\rangle_{k+1} \otimes ... \otimes |1\rangle_n, \tag{9}$$

where $Z_{k+1}, ..., Z_n$ is the stabilizer generator of the trivial code [34]. Therefore, only logical states in the codespace will be decoded into the trivial code state with all nondata qubits remaining in the $|1\rangle$. A multi-qubit control phase gate will therefore gain the necessary information to detect if the logical

state has left the codespace. For arbitrary stabilizer codes, the general decomposition of $U_{\bar{P}}$ can be found by Gottesman's algorithm [35].

IV. ERROR DETECTION WITH ICEBERG CODE

$[[2m + 2, 2m, 2]]$ Iceberg code is named by its planar connectivity requirement [Fig. 3(a)] and has stabilizer generators $\{X^{\otimes 2m+2}, Z^{\otimes 2m+2}\}$. Although this kind of distance-2 codes cannot correct any errors but only detect a single error, the large encoding rate and loose connectivity requirements render great practical interest in near-term algorithm application [26,36,37]. To demonstrate the performance of the GSM method, we numerically simulate the state in density matrix formalism using *qiskit.quantum_info* [38]. We take the noise parameters from the latest *Sycamore* processor [4] (with tunable couplers) and the *ibm_brisbane* processor [30] (with fixed-frequency couplers). We consider initialization noise, gate noise, idling decoherence, and readout assignment error. Particularly, we model the two-qubit gate noise as a combination of a depolarizing channel and an idling decoherence channel with duration equal to the gate time to fully reflect the effect of gate time on the performance. We discuss our simulation in more details in the Appendix A. Although we only simulate the case when $m = 1$ and $m = 2$, we expect larger Iceberg code should exhibit similar numerical behavior as the number of stabilizer generators does not increase with the code size.

We first use noisy gates to prepare the physical qubits in the GHZ state

$$\frac{1}{2}(|0\rangle^{\otimes 2m+2} + |1\rangle^{\otimes 2m+2}), \tag{10}$$

which is the logical $|\overline{+}\rangle^{\otimes 2}$ for $[[4, 2, 2]]$ code and $|\overline{+}\rangle^{\otimes 4}$ for $[[6, 4, 2]]$ code up to local transversal Hadamard. Next, we apply a varying period of idling noise on every physical qubit to mimic realistic noise. We then separately use the canonical SM or the GSM method to detect the error and post-select
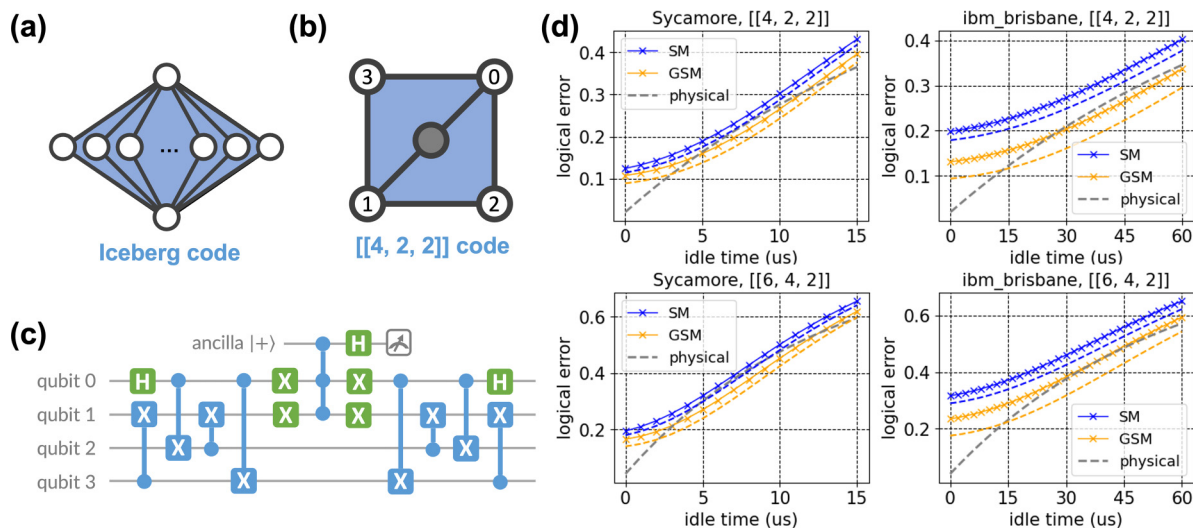

FIG. 3. (a) Planar connectivity structure of Iceberg codes. All physical qubits are placed in an iceberg-like shape and the bottom/up qubit is connected to every other qubit. (b) The [[4, 2, 2]] code with the required ancilla connectivity for GSM. (c) Decomposed circuit for GSM on the [[4, 2, 2]] code. (d) Numerical simulation of four-qubit and six-qubit Iceberg code, with noise parameters of *Sycamore* [4] and *ibm_brisbane* processor [30]. The dashed curve indicates error rate when the gate control is perfect and all gate noise is gate-induced idling noise.
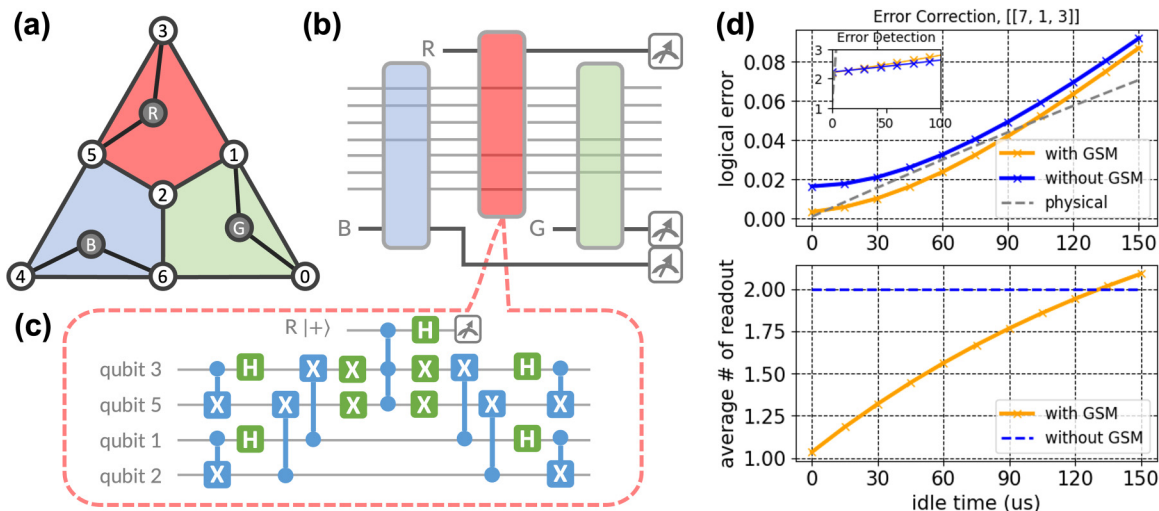
FIG. 4. (a) Planar connectivity structure of Steane code for the GSM method. On each color plaquette, the ancilla qubit is only required to connect two neighboring physical qubits on the edge. (b) Circuit scheme for three-shot GSM on Steane code. The check on each plaquette is performed sequentially but the ancilla qubits are measured in the same time. (c) Detailed implementation for plaquette check. (d) Logical error rate under idling noise after performing QEC cycle with (without) GSM and the corresponding number of average readout required. The inset figure is the logical error rate after just performing postselection with quantum error detection (y axis label in the unit of 0.001).

the state without triggered syndrome. We adopt the sandwich implementation of the control projector gate and the decomposed circuit for the four-qubit code is shown in Fig. 3(c), which requires the implementation of physical CCZ gates or further decomposition. This implementation suits better for quantum hardware with less flexible qubit connectivity. The general implementation of the GSM method on the Iceberg code is discussed in Appendix E. We plot the logical error rate against the idling time in Fig. 3(d) as well as the raw physical state prepared in $|+\rangle^{\otimes 2m}$ under idling noise for reference. We also plot the logical error rate when the gate noise is fully dominated by gate-induced decoherence as the dashed curve with the same colors.

For all four simulations shown in Fig. 3(d), the GSM method outperforms the canonical SM method by a significant gap. The amount of error rate improvement for the GSM method is barely affected by the idling time applied to the input logical state. In the case of state-of-the-art quantum hardware, the canonical SM method does not even have a better performance than the unencoded physical state, i.e., in a regime where the logical error rate is lower than the raw error rate. Nevertheless, the GSM method can achieve break even and outperform the unencoded physical state after a certain idling time. The advantage of GSM is even larger when the control is perfect and gate noise is fully dominated by the gate-induced decoherence. As the gate control techniques are consistently improving while the gate time is limited by physical constraints in many hardware implementations, such as the state leakage in superconducting qubits [1] or the operating laser power [13] for neutral atoms, we anticipate the gate noise in the future quantum hardware to be more dominated by gate-induced decoherence and our protocol could therefore bring further advantage over the canonical protocol.

V. ERROR CORRECTION WITH STEANE CODE

Although quantum error detection can bring benefits for very near-term scenario, the postselection routine introduces sampling overhead and quantum error correction is needed from a scalable viewpoint. To demonstrate the advantage of the GSM method for quantum error correction, we choose the [[7, 1, 3]] Steane code to numerically compare the logical performance of the GSM with canonical SM method. The Steane code is the smallest two-dimensional color code that is able to correct a single error, and has weight-four stabilizer $XXXX$ and $ZZZZ$ on each of the three color plaquettes shown in Fig. 4(a). Rather than using the one-shot GSM method that projects the logical state onto the codespace with a single readout, we choose to use the three-shot GSM to alleviate the connectivity requirement for planar layout of physical qubits. Specifically, we use three ancilla qubits, each to measure the projector of stabilizer $XXXX$ and $ZZZZ$ on its color plaquette. As all the three ancilla can be measured out simultaneously, the three-shot GSM method can still reduce the readout time as the canonical SM require two readouts on each ancilla qubit. In principle, we can use the circuit implementation for four-qubit Iceberg code shown in Fig. 3(c) for each plaquette check as the stabilizer set is the same. However, we use an alternative implementation [Fig. 4(c)] here to avoid the excessive connection on any single qubits such that each physical qubit is connected to at most four other physical qubits. In such a way, our GSM method should be exploited in state-of-the-art superconducting hardware without any further physical modification.

We use a projective noise parameter set for Steane code simulation because the state-of-the-art quantum hardware is still too noisy for logical qubits to outperform physical qubits see Appendix C. We initialize the logical qubit in $|\mp\rangle$ with noisy circuits, apply idling noise to every physical qubit, and

perform quantum error correction on the noisy logical state. We compare the performance of the QEC cycle with GSM [Fig. 1(c)] and without GSM [Fig. 1(a)], and plot the logical error rate and the average number of readout in Fig. 4(d). More simulation details can be found in the Method section as well.

With the help of the GSM method, the logical error rate after a QEC cycle is significantly reduced when the applied idling noise is low. We attribute the advantage of GSM to the reduced average time of readout: When there is little idling noise, the input state is nearly perfect and the precheck routine can assure the correctness of the input state with only a single-shot readout, leading the average number of readout close to one. In contrast, the canonical SM method always needs to measure twice and the readouts itself introduce additional noise to the logical state. When the idling noise increases and the input state becomes noisier, the average time of readout is increased as the GSM routine has more chance to detect an error and trigger the following correction process. However, even when the average time of readout is the same for QEC with or without GSM, we still see nonzero advantage in the QEC process with GSM. We attribute this advantage to the higher robustness of the GSM method on false-positive readout assignment errors: When the logical state is in the codespace but the readout outcome wrongly indicates there is an error in the detection stage, the QEC cycle with the GSM method will not be affected as the following correction routine will not detect any error and assure that the logical state is in the codespace.

VI. DISCUSSION

A natural question arise from our work is: Does there exist a range where the GSM method does not bring advantage on the logical fidelity? As the GSM method trades additional gate overhead for reduced readout overhead, the ratio between readout time and gate time can be used as the criterion. We numerically study the effect of readout time and gate time on the relative performance of the SM method and the GSM method in the supplemental material in Appendix D and find out the parameter of practical superconducting quantum hardware falls in the regime where the GSM method is advantageous. Therefore, unless there is a technical breakthrough that shortened the physical readout time drastically, we believe the GSM method will bring practical advantage for near-term quantum applications.

Besides using the GSM method as a precheck routine, fault-tolerant quantum computing can also benefit from GSM by exploit concatenated codes and treat detected physical errors as logical erasure errors [39], which does not require explicit quantum error correction in the inner code but only need to correct logical erasure in the outer loss-tolerant codes (see Appendix G). In such a case, our GSM method can be exploited to improve the robustness against readout-induced error.

Notably, as the $e^{i\pi\bar{P}}$ gate used in the GSM method needs to be decomposed into elementary gates, local errors might propagate to other physical qubits and produce correlated errors. However, as long as readout-induced idling error, which is a global noise channel, still serves as the main noise contributor in the near-term scenario, the correlated errors would not weigh over the readout-induced errors and pose a critical threat to the validness of our proposed method. Moreover, we may also use the idea of Shor-style syndrome measurement, which rely on pre-entangled ancilla qubits and native multi-qubit Pauli gates to achieve gate-level fault tolerance at the cost of more ancilla qubits and time overhead (see Appendix F).

Although we only simulate the performance of the GSM method using noise parameters from superconducting hardware, the GSM method can generally be applied to various kinds of quantum hardware where mid-circuit readouts are error prone. For neutral atoms, state-of-the-art readout time is on the scale of tens of milliseconds and the best reachable coherence time is on the scale of seconds [8,40]. In particular, mid-circuit measurements for neutral atoms rely on coherently transporting the atoms to a separate area as shown in [8], or need to shelve the data qubits to magnetic-sensitive states with reduced coherence time [41]. Our protocol can therefore be employed to provide advantages for scalable neutral atoms quantum computers as well.

APPENDIX A: SIMULATION DETAILS

All of our numerical results are obtained from simulation using *qiskit.quantum_info* [38]. The quantum states are simulated in density matrix formalism and all error channels are implemented as Kraus superoperators. All physical qubits are prepared in $(1 - \epsilon_{\text{init}})|0\rangle\langle 0| + \epsilon_{\text{init}}|1\rangle\langle 1|$ where ϵ_{init} is the initialization error. For single-qubit (1Q) gates, we only consider the control error and model the noise by following a depolarizing channel after the ideal gate operation. For two-qubit (2Q) gates, we consider both the control error and the gate-induced decoherence. We follow a two-qubit depolarizing channel after the ideal 2Q gate operation, and a global idling channel to all physical qubits after a layer of 2Q gates. In such a way, we ensure the total infidelity of a 2Q gate under control noise and the idling noise equals to the median error rate reported in [4] and [30]. For the Toffoli gate used in the GSM method, we model the noise by follow a three-qubit depolarizing channel with double error rate of 2Q gate and double gate-induced idling duration. Although most large-scale quantum hardware by far does not support native Toffoli gates and require further decomposition into five two-qubit gates, we are taking an optimistic consideration as native multi-qubit gates have been demonstrated on various kinds of quantum hardware [13,42]. For noisy readout, we model it by inserting an idling channel whose duration is equal to readout time before measurement. We also considered the readout (assignment) error that read $|0\rangle$ as $|1\rangle$ and vice versa. The specific noise parameters can be found in Table I.

TABLE I. Noise parameter for numerical simulation.

	Sycamore	*ibm_brisbane*	Projective
T_1 (μs)	20	217	1000
T_2 (μs)	30	130	1000
readout time(ns)	660[a]	4000	200
1Q gate error	1e-3	2.2e-4	0
2Q gate time (ns)	34	600	20
2Q gate error	3e-3	7.5e-3	1e-4
initialize error	1e-2	1e-2	1e-3
readout error	2e-2	1e-2	1e-3

[a]Including time for qubit resetting.

For the Iceberg code simulation, we initialize the logical state by preparing a multi-qubit GHZ state. We start with a physical qubit in $|+\rangle$ and all others in $|0\rangle$ and apply CNOT gates sequentially. The depth of initialization for four(six)-qubit code is 2(3) and can be achieved with the iceberg-like connectivity. We assume only one available ancilla, which connects to every other physical qubit in the SM method but only require connection to two physical qubits in the GSM method. The gate depth of a cycle of the GSM method is 4(8) plus a Toffoli depth for the 4(6)-qubit Iceberg code. Comparatively, the depth for the SM method is 8(12).

For the Steane code simulation, we prepare the logical state by applying nine CZ gates (depth 3) on physical qubits all prepared in $|+\rangle$ and Hadamard gates on four of them [8]. We use three ancilla qubits, one for each color plaquette. In the QEC cycle without the GSM method, we first perform the X check on each plaquette, then the Z check on each plaquette. In the QEC cycle with the GSM method, we apply the partial projector gate on each ancilla sequentially [Fig. 4(b)] before reading out all three ancilla qubits. The gate depth for the GSM method is 12 plus 3 Toffoli depth. In contrast, the canonical SM has 8 depth of 2Q gates.

APPENDIX B: *m*-SHOT GENERALIZED SYNDROME MEASUREMENT WITH SUBPROJECTORS

In the main text, we introduced the generalized syndrome measurement (GSM), which only requires a single readout to detect the error. The GSM method requires the ability to implement controlled projector gates $e^{i\pi\bar{P}}$, which could bring large computational overhead when decomposed into elementary gates and the connectivity requirement could be very strict. Here we introduce the modified method by exploiting a bit more ancilla qubits, which could help reach a smooth trade-off between the gate complexity/connectivity and the readout complexity in the GSM method.

For code Q with stabilizer generator $\{S_i\}$, the general projector for one-shot GSM method is

$$\bar{P} = \prod_i^k P_i = \frac{1}{2^k}\prod_i^k(I + S_i) = \frac{1}{K}\sum_i^K M_i. \quad (B1)$$

As we showed in the main text, we can use a control-$e^{i\pi\bar{P}}$ gate and a single-shot measurement on the ancilla qubit to act the projector on the noisy state nondeterministically. Instead of using only one projector, we can group several subprojectors
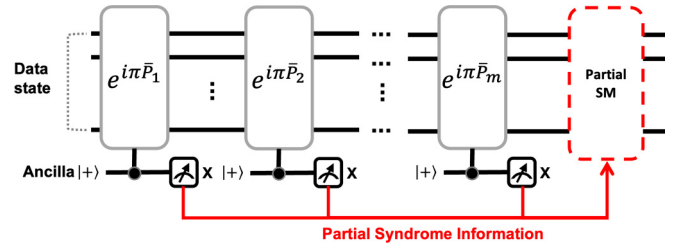

FIG. 5. Circuit scheme for *m*-shot generalized syndrome measurement.

$\bar{P}_i$ such that their product is still $\bar{P}$,

$$\bar{P} = \prod_i^m \bar{P}_i, \quad (B2)$$

where

$$\bar{P}_i = \prod_{\{j\}_i} P_j \quad (B3)$$

and $\{j\}_i$ is an index set associated with each subprojector and $\bigcup_{i=1}^m \{j\}_i = \{x \in \mathbb{Z} | 0 \leqslant x \leqslant k\}$. As we have m subprojectors, we need m ($1 \leqslant m \leqslant k$) readouts for partial syndrome extraction here. We can use the circuit show in Fig. 5 to extract the partial syndrome in a QEC cycle.

As a convenience choice, we can choose to group stabilizers pairwise as $\{S_{2i-1}, S_{2i}\}$, which allows us to define the two-element subprojectors $\bar{P}_i^2$,

$$\bar{P}_i^2 = P_{2i-1}P_{2i} = \frac{1}{4}(I + S_{2i-1})(I + S_{2i}), \quad (B4)$$

and it is easy to see

$$\bar{P} = \prod_i^{k/2} \bar{P}_i^2. \quad (B5)$$

In such a way, we can halve the amount of readouts while the implementation of the projector gates $e^{i\pi\bar{P}_i^2}$ is still not too complicated. For instance, the [[5, 1, 3]] code has four stabilizer generators

$$S_1 = ZXXZI \quad S_2 = IZXXZ$$
$$S_3 = ZIZXX \quad S_4 = XZIZX, \quad (B6)$$

and we can just use a single readout to measure the projector

$$\bar{P} = \frac{1}{16}(I + ZXXZI)(I + IZXXZ)$$
$$\times (I + ZIZXX)(I + XZIZX). \quad (B7)$$

This would require 18 two-qubit gates (depth 12) and a five-qubit Toffoli gate in total to implement the projector gate. Rather if we measure the following two subprojectors

$$\bar{P}_1 = \frac{1}{4}(I + ZXXZI)(I + IZXXZ)$$
$$\bar{P}_2 = \frac{1}{4}(I + ZIZXX)(I + XZIZX), \quad (B8)$$

we only need 12 two-qubit gates (depth 6) and a Toffoli gate for each subprojector gate and the overall gate overhead is 24 two-qubit gates (depth 12) and two Toffoli gates.

Importantly, the partial syndrome information extracted with the GSM method can be further used in the correction
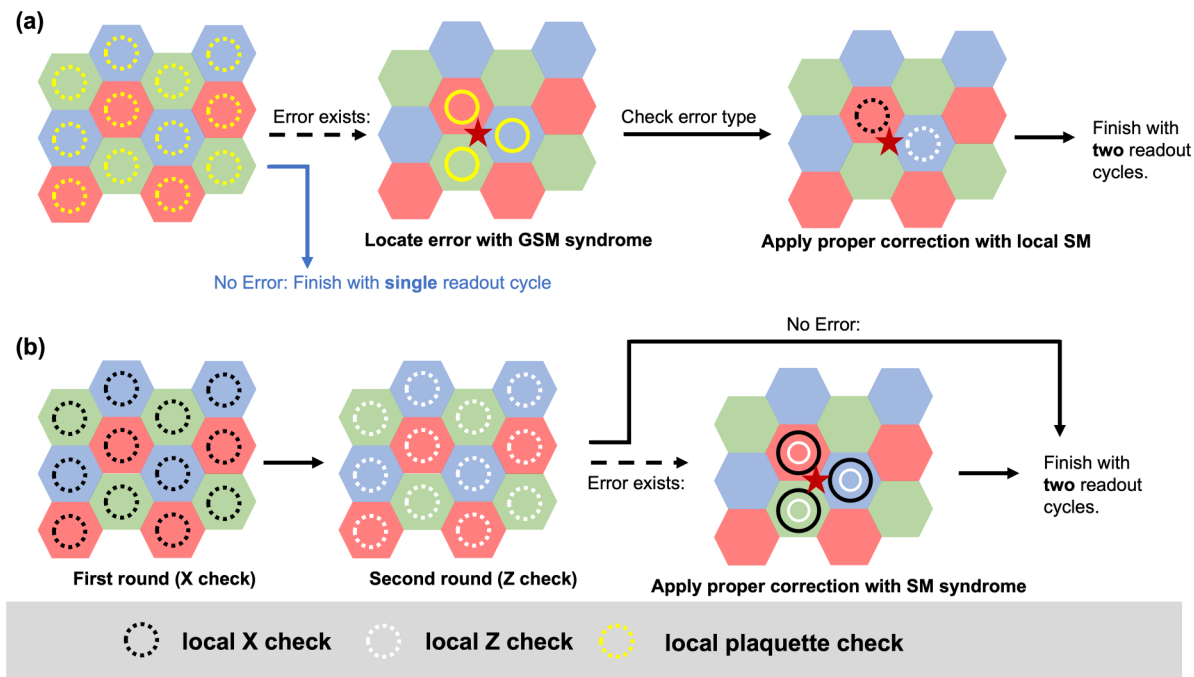
FIG. 6. Adoption of the GSM method on scalable two-dimensional color codes. (a) A QEC cycle with the GSM method. Yellow dashed circles represent for the local GSM routine on each plaquette. When error rate is low, we will mostly end with just a single-shot readout time. We follow with local stabilizer check when the location of error(s) is detected to correct the error(s), and the time consumed is two-shot readout time. (b) A canonical QEC cycle without the GSM method. Whether an error is detected or not, the time required is always two-shot readout time.

stage in the QEC cycle. We only need to perform canonical syndrome measurement for stabilizers associated with triggered subprojectors, and stabilizers associated with untriggered subprojectors does not need to be measured again. Taking the two-shot GSM method on the $[[5, 1, 3]]$ code as an example: If only the second subprojector $\bar{P}_2$ is triggered, we only need to perform canonical syndrome measurement for stabilizer $ZIZXX$ and $XZIZX$, and there is no need to measure $ZXXZI$ and $IZXXZ$ again as we already assured that the logical state is in the space of subprojector $\bar{P}_1$.

For another example, the two-dimensional color code [43] has stabilizer $\{X^{\otimes n}, Z^{\otimes n}\}$ supported on each n-edge plaquette. We can use a subprojector associated with each plaquette

$$\bar{P}_{plaq} = \tfrac{1}{4}(I + X^{\otimes n})(I + Z^{\otimes n}) \tag{B9}$$

to extract partial syndrome information (Fig. 6). For this kind of code, a QEC cycle with our GSM method only takes a single readout cycle when no error occurs, and at most two readout cycles in total when at least an error occurs. To be compared with, a canonical QEC cycle without the GSM method always takes two readout cycles. Therefore, our GSM method can always reduce the time of readout for two-dimensional color code, providing a scalable option for fault-tolerant quantum computing.

In general, there is no restriction on how to group the stabilizer generators into subprojectors, and the subprojectors can be chosen to mostly suit the connectivity and practical limitation of quantum hardware. The number of readouts can be chosen smoothly from 1, corresponding to the one-shot GSM method proposed in the main text with the largest gate overhead (minimum readout overhead), to k, corresponding to

the canonical syndrome measurement with the minimum gate overhead (largest readout overhead).

APPENDIX C: STEANE CODE SIMULATION WITH STATE-OF-THE-ART NOISE PARAMETER

In the main text, we present numerical results for the GSM method on the Steane code with projective noise parameters. The main reason why we do not stick with state-of-the-art noise parameter is the current superconducting quantum hardware is still too noisy and not in the regime where logical qubit would outperform physical qubit, as shown in Fig. 7. The logical error rate after the encoding and decoding is much larger than the physical error rate, which makes QEC not really meaningful in such a regime. However, we still see some advantage of the GSM method over the canonical SM method when the input logical state does not suffer too long idling time. As longer idling noise is applied to the logical state, the average number of readout of a QEC cycle with the GSM method increases over two (canonical SM method) and the GSM method start to lose its advantage. As raw physical qubits dominates over the logical qubits, this regime does not post a critique to our proposed GSM method.

APPENDIX D: BENCHMARK THE ADVANTAGEOUS REGIME OF THE GSM METHOD

To understand the advantageous regime of the GSM method, we ramp both the readout time and the gate time for both the *Sycamore* and the *ibm_brisbane* setting within a practical range that might be reached in the future, while keeping
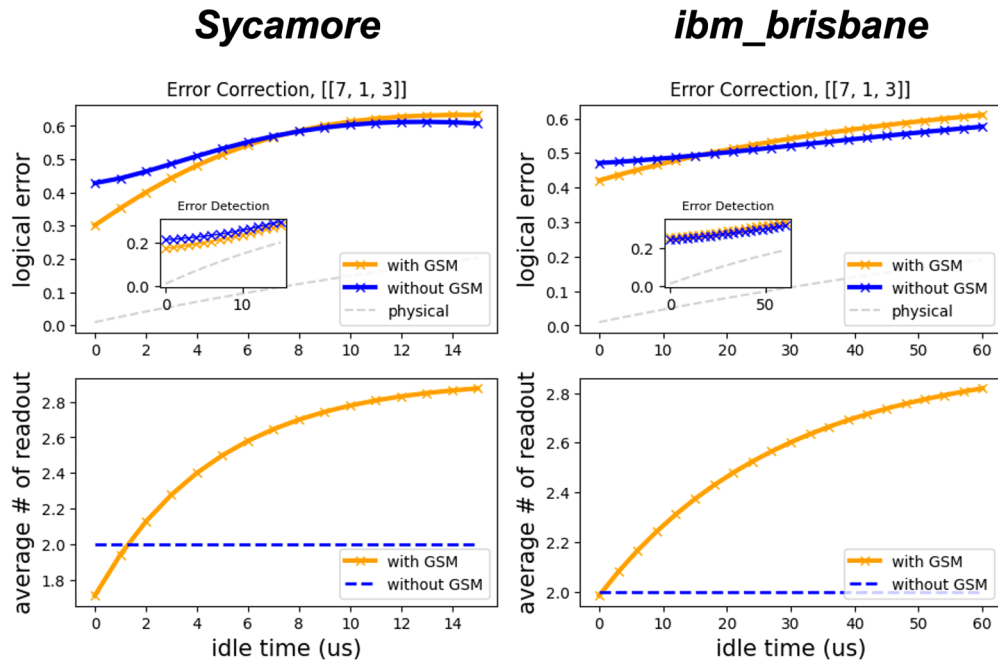
FIG. 7. Steane code simulation with state-of-the-art noise parameter.

other noise parameters fixed. We compare the performance for both the GSM and the canonical SM method and plotted the fidelity difference $F_{\text{gsm}} - F_{\text{sm}}$ regarding the gate time and the readout time. We simulated both the [[4, 2, 2]] code in Fig. 8 and the [[6, 4, 2]] code in Fig. 9 to explore the condition when the GSM method is advantageous. The advantageous regime

for the GSM(SM) method is plotted with red(blue) area. From the simulated results, we see the state-of-the-art parameter is in a range where the GSM method will beat the canonical SM method. Unless the readout duration is significantly reduced in the scalable quantum hardware, we will still stay in the advantageous regime for the GSM method.
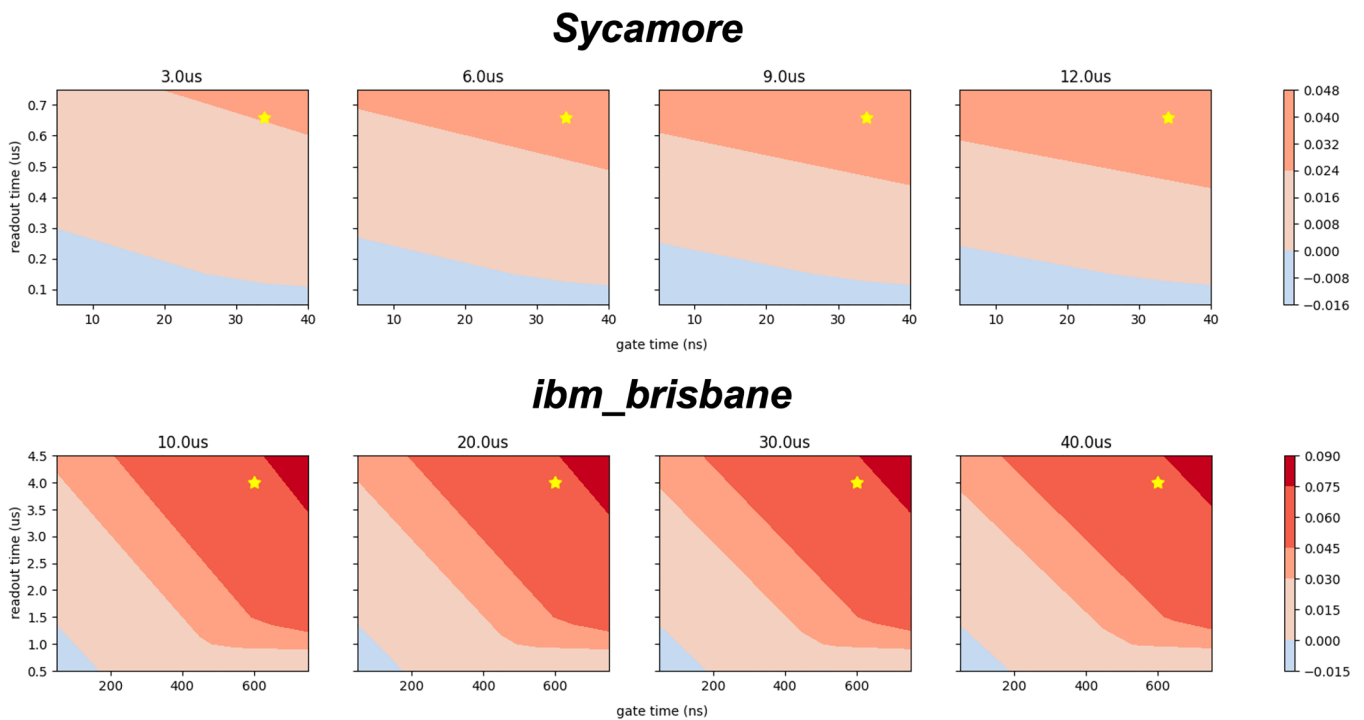

FIG. 8. Fidelity difference $F_{\text{gsm}} - F_{\text{sm}}$ with [[4, 2, 2]] code depending on various readout time and gate time when other noise parameters are fixed. The yellow stars denote where the state-of-the-art operation time is.
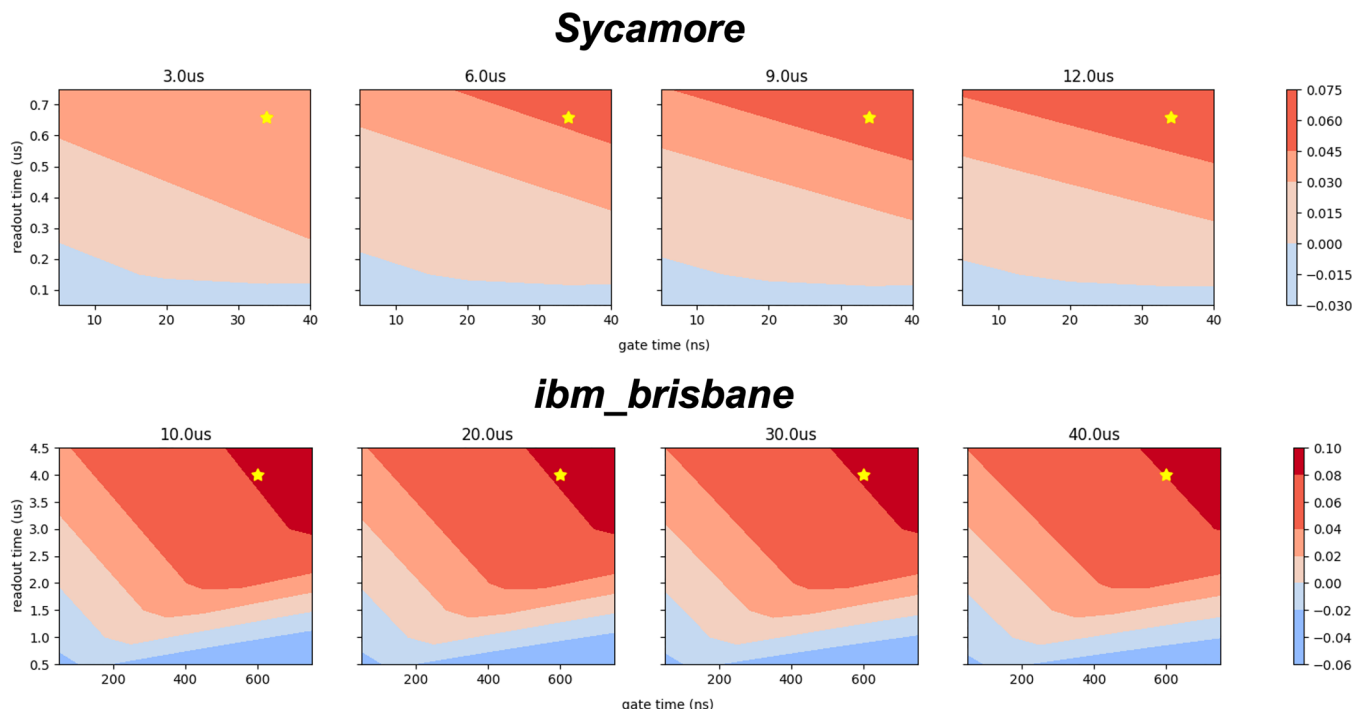
FIG. 9. Fidelity difference $F_{gsm} - F_{sm}$ with [[6, 4, 2]] code depending on various readout time and gate time when other noise parameters are fixed. The yellow stars denote where the state-of-the-art operation time is.

APPENDIX E: PROJECTOR GATE IMPLEMENTATION FOR ICEBERG CODE

Although the implementation of the projector gate is not unique, here we present an implementation for general Iceberg code in Fig. 10. In such an implementation, the 2Q gate depth is only $4m$. Comparatively, the overall 2Q depth of the canonical syndrome measurement is $4m + 4$.

APPENDIX F: FAULT-TOLERANCE IN ANALOG WITH SHOR-STYLE SYNDROME MEASUREMENT

In the long term, we have to consider the gate-level fault tolerance of our method to avoid the impact of correlated errors. To achieve this in our method, we require the quantum hardware to be able to implement native multi-qubit (control) Pauli phase gates and does not rely on decomposition into elementary gates. It is reasonable to make such an assumption as various quantum hardware has demonstrated potential for native multi-qubit gates [12,42]. We also demand that the

native global gates do not suffer from inner correlated errors. As such, we only need to consider single Pauli errors induced by the gate operations up to the main order.

By using the idea of Shor-style syndrome measurement, i.e., preparing entangled ancilla states, we can achieve a fault-tolerant circuit for our method. We plotted the GSM circuit for [4, 2, 2] code as an example in Fig. 11. The entangled ancilla states can prevent errors occurred in any single ancilla qubit. For errors occurred after the multi-qubit Pauli gates, they will either be detected by the following Pauli gates or commute with the following Pauli gates and will not be propagated to other qubits (Fig. 11).

Importantly, we note this fault tolerance cannot be achieved if we decompose the gate into elementary gates (two-qubit gates). That is because each Pauli phase gate is mediating global interaction and an error occurred after a decomposed gate could propagate to every qubit at the end of the Pauli phase gate. In early fault-tolerant setting, this may not be a
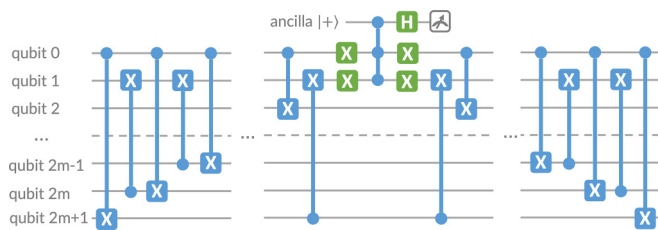

FIG. 10. The projector gate implementation of for [[2m + 2, 2m, 2]] Iceberg code.

FIG. 11. Generalized Syndrome Measurement in analog with Shor-style syndrome measurement for [[4, 2, 2]] code. Ancilla qubits (green) are prepared in the GHZ state $(|000\rangle + |111\rangle)/\sqrt{2}$ and each ancilla qubit controls a multi-qubit Pauli gate. An error occurred after the gates will either be detected (left) or commute with the following gates and will not propagate (right).
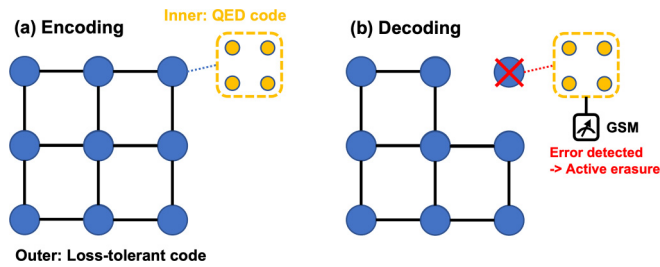
FIG. 12. Scheme for concatenated codes with logical erasure conversion. In such a scenario, the ability to detect the errors is sufficient to correct physical errors.

severe problem if the gain from the reduced readout-induced noise is good enough, as shown in our simulation results.

APPENDIX G: QUANTUM ERROR CORRECTION WITH LOGICAL ERASURE CONVERSION

We can exploit the benefit of concatenated quantum codes to correct the errors indirectly (Fig. 12). The idea is initially proposed in [39] for cluster state generations. The inner codes are set to be small-size quantum code used to detect the physical errors, and the outer codes are used to correct the loss (erasure) errors. Whenever we detect an error in the inner level, we actively treat it as a logical erasure error in the outer level. As outer codes are loss tolerant, we can restore the logical information in the presence of logical erasures and thus correct the physical errors indirectly. We do not need the syndrome information and correct the Pauli errors directly, and the GSM method can help reduce the readout-induced infidelity in such a scenario.

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[2] M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo *et al.*, Quantum approximate optimization of non-planar graph problems on a planar superconducting processor, Nat. Phys. **17**, 332 (2021).

[3] M. Kjaergaard, M. E. Schwartz, A. Greene, G. O. Samach, A. Bengtsson, M. O'Keeffe, C. M. McNally, J. Braumüller, D. K. Kim, P. Krantz, M. Marvian, A. Melville, B. M. Niedzielski, Y. Sung, R. Winik, J. Yoder, D. Rosenberg, K. Obenland, S. Lloyd, T. Orlando *et al.*, Demonstration of density matrix exponentiation using a superconducting quantum processor, Phys. Rev. X **12**, 011005 (2022).

[4] Google Quantum AI, R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, Nature (London) **614**, 676 (2023).

[5] Y. Zhao, Y. Ye, H.-L. Huang, Y. Zhang, D. Wu, H. Guan, Q. Zhu, Z. Wei, T. He, S. Cao, F. Chen, T.-H. Chung, H. Deng, D. Fan, M. Gong, C. Guo, S. Guo, L. Han, N. Li, S. Li *et al.*, Realization of an error-correcting surface code with superconducting qubits, Phys. Rev. Lett. **129**, 030501 (2022).

[6] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann *et al.*, Realizing repeated quantum error correction in a distance-three surface code, Nature (London) **605**, 669 (2022).

[7] J. F. Marques, B. M. Varbanov, M. S. Moreira, H. Ali, N. Muthusubramanian, C. Zachariadis, F. Battistel, M. Beekman, N. Haider, W. Vlothuizen *et al.*, Logical-qubit operations in an error-detecting surface code, Nat. Phys. **18**, 80 (2022).

[8] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner *et al.*, A quantum processor based on coherent transport of entangled atom arrays, Nature (London) **604**, 451 (2022).

[9] K. Singh, S. Anand, A. Pocklington, J. T. Kemp, and H. Bernien, Dual-element, two-dimensional atom array with continuous-mode operation, Phys. Rev. X **12**, 011040 (2022).

[10] T. M. Graham, Y. Song, J. Scott, C. Poole, L. Phuttitarn, K. Jooya, P. Eichler, X. Jiang, A. Marra, B. Grinkemeyer *et al.*, Multi-qubit entanglement and algorithms on a neutral-atom quantum computer, Nature (London) **604**, 457 (2022).

[11] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi *et al.*, Generation and manipulation of Schrödinger cat states in Rydberg atom arrays, Science **365**, 570 (2019).

[12] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, and M. D. Lukin, Parallel implementation of high-fidelity multiqubit gates with neutral atoms, Phys. Rev. Lett. **123**, 170503 (2019).

[13] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara *et al.*, High-fidelity parallel entangling gates on a neutral-atom quantum computer, Nature (London) **622**, 268 (2023).

[14] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature (London) **626**, 58 (2024).

[15] M. Ringbauer, M. Meth, L. Postler, R. Stricker, R. Blatt, P. Schindler, and T. Monz, A universal qudit quantum processor with trapped ions, Nat. Phys. **18**, 1053 (2022).

[16] P. T. Dumitrescu, J. G. Bohnet, J. P. Gaebler, A. Hankin, D. Hayes, A. Kumar, B. Neyenhuis, R. Vasseur, and A. C. Potter, Dynamical topological phase realized in a trapped-ion quantum simulator, Nature (London) **607**, 463 (2022).

[17] 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).

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

[19] J. Preskill, Reliable quantum computers, Proc. R. Soc. A **454**, 385 (1998).

[20] V. Nebendahl, H. Häffner, and C. F. Roos, Optimal control of entangling operations for trapped-ion quantum computing, Phys. Rev. A **79**, 012312 (2009).

[21] G. A. Paz-Silva, G. K. Brennen, and J. Twamley, Fault tolerance with noisy and slow measurements and preparation, Phys. Rev. Lett. **105**, 100501 (2010).

[22] D. Crow, R. Joynt, and M. Saffman, Improved error thresholds for measurement-free error correction, Phys. Rev. Lett. **117**, 130503 (2016).

[23] S. Heußen, D. F. Locher, and M. Müller, Measurement-free fault-tolerant quantum error correction in near-term devices, PRX Quantum **5**, 010333 (2024).

[24] Y. Zhou, Z. Zhang, Z. Yin, S. Huai, X. Gu, X. Xu, J. Allcock, F. Liu, G. Xi, Q. Yu *et al.*, Rapid and unconditional parametric reset protocol for tunable superconducting qubits, Nat. Commun. **12**, 5924 (2021).

[25] Y. Sunada, S. Kono, J. Ilves, S. Tamate, T. Sugiyama, Y. Tabuchi, and Y. Nakamura, Fast readout and reset of a superconducting qubit coupled to a resonator with an intrinsic Purcell filter, Phys. Rev. Appl. **17**, 044016 (2022).

[26] C. N. Self, M. Benedetti, and D. Amaro, Protecting expressive circuits with a quantum error detection code, Nat. Phys. **20**, 219 (2024).

[27] A. Steane, Multiple-particle interference and quantum error correction, Proc. R. Soc. A **452**, 2551 (1997).

[28] N. Delfosse and B. W. Reichardt, Short Shor-style syndrome sequences, arXiv:2008.05051.

[29] T. Tansuwannont, B. Pato, and K. R. Brown, Adaptive syndrome measurements for Shor-style error correction, Quantum **7**, 1075 (2023).

[30] IBM Quantum Experience, https://quantum.ibm.com/.

[31] J. v. d. Wetering, Constructing quantum circuits with global gates, New J. Phys. **23**, 043015 (2021).

[32] S. Jandura and G. Pupillo, Time-optimal two- and three-qubit gates for Rydberg atoms, Quantum **6**, 712 (2022).

[33] L. Clinton, T. Cubitt, B. Flynn, F. M. Gambetta, J. Klassen, A. Montanaro, S. Piddock, R. A. Santos, and E. Sheridan, Towards near-term quantum simulation of materials, Nat. Commun. **15**, 211 (2024).

[34] R. Cleve and D. Gottesman, Efficient computations of encodings for quantum error correction, Phys. Rev. A **56**, 76 (1997).

[35] D. Gottesman, Stabilizer codes and quantum error correction, arXiv:quant-ph/9705052.

[36] K. Yamamoto, S. Duffield, Y. Kikuchi, and D. Muñoz Ramo, Demonstrating Bayesian quantum phase estimation with quantum error detection, Phys. Rev. Res. **6**, 013221 (2024).

[37] M. Urbanek, B. Nachman, and W. A. de Jong, Error detection on quantum computers improving the accuracy of chemical calculations, Phys. Rev. A **102**, 022427 (2020).

[38] Qiskit contributors, Qiskit: An open-source framework for quantum computing, https://doi.org/10.5281/zenodo.2573505.

[39] Z. Li, I. Kim, and P. Hayden, Concatenation schemes for topological fault-tolerant quantum error correction, Quantum **7**, 1089 (2023).

[40] K. Singh, C. E. Bradley, S. Anand, V. Ramesh, R. White, and H. Bernien, Mid-circuit correction of correlated phase errors using an array of spectator qubits, Science **380**, 1265 (2023).

[41] T. M. Graham, L. Phuttitarn, R. Chinnarasu, Y. Song, C. Poole, K. Jooya, J. Scott, A. Scott, P. Eichler, and M. Saffman, Mid-circuit measurements on a single-species neutral alkali atom quantum processor, Phys. Rev. X **13**, 041051 (2023).

[42] Y. Kim, A. Morvan, L. B. Nguyen, R. K. Naik, C. Jünger, L. Chen, J. M. Kreikebaum, D. I. Santiago, and I. Siddiqi, High-fidelity three-qubit iToffoli gate for fixed-frequency superconducting qubits, Nat. Phys. **18**, 783 (2022).

[43] A. G. Fowler, Two-dimensional color-code quantum computation, Phys. Rev. A **83**, 042310 (2011).