

Fault-Tolerant Quantum Computation with Constant Overhead for General Noise

Matthias Christandl^{1,*}, Omar Fawzi^{2,†} and Ashutosh Goswami^{1,‡}

¹*Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark*

²*Laboratoire de l'Informatique du Parallélisme (LIP), Université de Lyon, Inria, École normale supérieure (ENS) de Lyon, Université Claude Bernard Lyon (UCBL), 69364 Lyon Cedex 07, France*



(Received 3 May 2025; revised 8 September 2025; accepted 18 September 2025; published 13 November 2025)

Fault-tolerant quantum computation traditionally incurs substantial resource overhead, with both qubit and time overheads scaling polylogarithmically with the size of the computation. While prior work by Gottesman has shown that constant qubit overhead is achievable under stochastic noise using quantum low-density parity-check (QLDPC) codes, it has remained an open question whether similar guarantees hold under more general, nonstochastic noise models. In this work, we address this question by considering a general circuit-level noise model defined via the diamond norm, which captures both stochastic and nonstochastic noise, including coherent and amplitude damping noise. We prove that constant qubit overhead fault-tolerant quantum computation is achievable in this general setting, using QLDPC codes with constant rate and linear minimum distance. To establish our result, we develop a fault-tolerant error correction scheme and a method for implementing logic gates under general circuit noise. These results extend the theoretical foundations of fault-tolerant quantum computation and offer new directions for fault-tolerant architectures under realistic noise models.

DOI: [10.1103/k4cm-pp9p](https://doi.org/10.1103/k4cm-pp9p)

I. INTRODUCTION

Fault tolerance is essential for large-scale quantum computations due to the inherently noisy nature of quantum systems [1,2]. A leading approach to fault-tolerant quantum computation is based on active error correction, where logical qubits encoded in the code space of an error-correcting code are created using many noisy physical qubits. Using fault-tolerant procedures, one can realize logic gates in the code space of the error-correcting code. The celebrated threshold theorem states that by doing active error correction, it is possible to realize arbitrarily long finite quantum computations with classical input and output, given that the noise rate in the quantum circuit is below a threshold value [3–5]. Fault tolerance, however, comes at a price of increased qubit count compared to the noiseless protocol. More precisely, to realize a quantum

circuit, one replaces it with a fault-tolerant quantum circuit, which operates on a much larger number of qubits. Therefore, it is of interest from a practical standpoint to minimize the qubit overhead, i.e., the ratio of the numbers of qubits in the fault-tolerant circuit and the original circuit.

Fault-tolerant quantum computing based on concatenated codes, which were the first error-correcting codes used to prove the threshold theorem, has a polylogarithmic asymptotic qubit overhead [3–5], meaning that the ratio of physical to logical qubits grows polylogarithmically with the size of the quantum computation. This naturally raises the question whether fault-tolerant quantum computation is possible with a qubit overhead that remains constant with the problem size. For stochastic noise, this question has been partially settled in Ref. [6], where a construction of constant-overhead fault-tolerant quantum computation has been provided using quantum low-density parity-check (QLDPC) codes. The construction in Ref. [6] was based on the assumption that a family of QLDPC codes with certain properties exists. More precisely, it supposes that there exists a family of constant-rate QLDPC codes that can suppress random errors with noisy syndrome extraction and that furthermore have an efficient classical decoding algorithm. It has then been shown in Ref. [7] that quantum expander codes satisfy all the requirements needed in Gottesman's construction; hence, establishing constant-overhead fault tolerance. It

*Contact author: christandl@math.ku.dk

†Contact author: omar.fawzi@ens-lyon.fr

‡Contact author: akg@math.ku.dk

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.

is important to note that Gottesman's construction leads to an increased overhead in depth due to a sequentialization step requiring that only a constant number of logic gates are applied at each time step. In particular, the overhead in depth is polynomial. In Ref. [8], a different approach based on concatenated codes has been used to show constant-overhead fault-tolerant quantum computation, where the time overhead is quasipolylogarithmic [9]. These results [6–8] have generated significant interest in constant-overhead quantum computation, particularly those based on QLDPC codes. Subsequent works have focused on optimizing these constant-overhead constructions for practical implementation [10–14].

However, all the previous works on constant-overhead fault-tolerant quantum computation [6–8,10–14] consider *stochastic* circuit noise (e.g., Pauli and local stochastic noise), leaving the question open whether it is possible to achieve constant qubit overhead under more realistic nonstochastic circuit noise (e.g., coherent and amplitude damping noise). Moreover, since the threshold theorem based on concatenated codes also applies to nonstochastic circuit noise [3,5], it is natural to ask whether such generalizations also hold for constant-overhead constructions.

In this work, we consider a general circuit noise, which includes both coherent noise and stochastic noise, as special cases. In particular, general noise with parameter δ replaces each gate g in the circuit by an arbitrary quantum channel $\tilde{g}$, which is δ -close to g in diamond norm, i.e., $\|g - \tilde{g}\|_\diamond \leq \delta$ [5]. We show that constant overhead can be achieved for general circuit noise using QLDPC codes that have constant rate and linear minimum distance. It is worth noting that linear minimum distance is not a requirement for stochastic circuit noise [6,7]. However, recent breakthroughs have demonstrated the existence of such QLDPC codes [15,16]; therefore, our results do not rely on any unproven assumptions.

To establish constant-overhead fault-tolerant quantum computation for general circuit noise, we revisit the construction from Refs. [6,7], and develop necessary tools

and generalizations to establish constant-overhead fault-tolerant quantum computation for general circuit noise. In particular, the following two points are necessary for constructions in Refs. [6,7]:

- (1) Fault-tolerant error correction for QLPDC codes can be done by measuring stabilizer generators of the code and performing error correction based on the measurement outcomes (syndrome).
- (2) The logic gates can be implemented using gate teleportation, where the required ancilla state in the code space can be prepared using concatenated codes.

We generalize the above two points for general circuit noise, considering QLDPC codes with constant overhead and linear minimum distance.

For the first point, we consider parallel error correction of several code blocks of QLDPC codes and analyze the syndrome extraction circuit for each code block under general circuit noise.

We divide the syndrome extraction Φ_{syn} circuit into three parts: (1) initialization of ancilla qubits in a Pauli- X or $-Z$ basis state; (2) application of controlled-NOT (CNOT) or controlled- Z (CZ) gates to entangle data qubits with ancilla qubits; and (3) measurement of the ancilla qubits in the Pauli- X or $-Z$ basis (see also Fig. 5 below). We show that the errors that happen during Φ_{syn} can be pushed just before the last layer of the circuit, where measurements are applied. Note that errors multiply as they are pushed since they can propagate through two qubit gates. However, error propagation is limited since Φ_{syn} has constant depth (with respect to the code length) for QLDPC codes, implying that the future light cone of any gate only contains a constant number of qubits (see Fig. 1 for an illustration of the future light cone and, below, Lemma 2, where a detailed analysis of Φ_{syn} under general circuit noise is given).

We then combine our analysis of the syndrome extraction circuit with the single-shot classical decoder introduced in Ref. [17] to show that the parallel error correction of several code blocks controls the errors in each code block, so that the computation can be sustained for an arbitrarily long time (see Theorem 5 for more details). It is worth emphasizing that parallel error correction is necessary since in our construction logical qubits are encoded in many smaller code blocks of QLDPC codes and not just in one big code block, similarly to Refs. [6,7] (see also Fig. 2). We note that, in this regard, our analysis improves on previous works [6,7], where error correction of only one code block has been considered, thereby ignoring correlations that arise between different blocks.

For the second point, we construct fault-tolerant logic gates by combining a fault-tolerant state preparation method for general circuit noise given in Ref. [18],

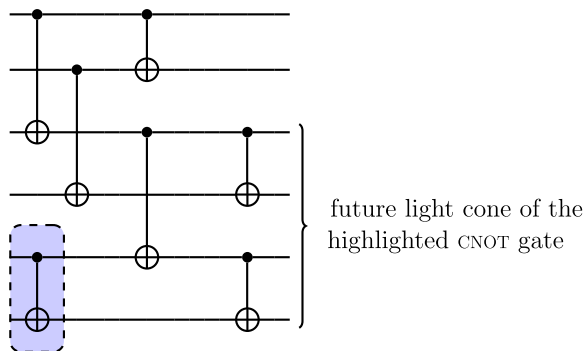


FIG. 1. The future light cone of the highlighted CNOT gate, i.e., the set of qubits that are affected by the CNOT gate.

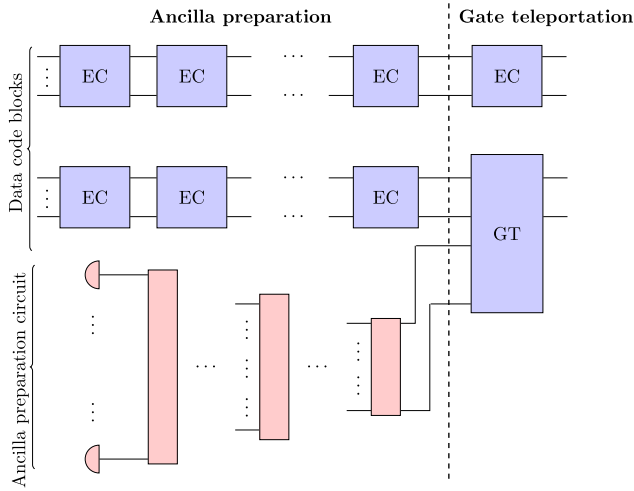


FIG. 2. The realization of a logic gate in the fault-tolerant circuit $\bar{\Phi}$ corresponding to Φ , according to Theorem 1. The blocks corresponding to EC (error correction) and GT (gate teleportation) represent code blocks of the QLDPC codes in $\bar{\Phi}$. The size of each block is $m = O(\sqrt{x})$, where x is the number of qubits in Φ . The logic gate is realized in two steps, namely, ancilla preparation and gate teleportation steps. During the ancilla preparation, a code state is prepared, using the state preparation method from Theorem 2. Note that the *decreasing size* of the layers in the ancilla preparation circuit reflects the fact that extra qubits are needed during ancilla preparation. The qubit overhead in the ancilla preparation circuit alone is in fact not constant (it is polylogarithmic). While the ancilla state is prepared, repeated rounds of error correction are applied on the data code blocks. During the gate teleportation, the prepared ancilla state is transferred to the specified code block, and the logic gate is realized using gate teleportation. On the remaining code blocks, error correction is applied.

together with gate teleportation. By integrating our generalizations of fault-tolerant error correction and logic gates within the constructions of Refs. [6,7], we establish the following theorem.

Theorem 1 (Informal, see Theorem 3 below for details). Consider a quantum circuit Φ , with classical input and output, working on x qubits, and of size (i.e., total number of locations) $\text{poly}(x)$. For any asymptotic overhead $\alpha > 1$, there exists a constant error rate $\delta_{\text{th}} > 0$, and a quantum circuit $\bar{\Phi}$, with the same input and output systems as Φ , and working on $x' = \alpha x + o(x)$ qubits, such that any noisy version $[\bar{\Phi}]_{\delta}$, under general circuit noise with parameter $\delta \leq \delta_{\text{th}}$, simulates Φ with a vanishing error probability as x goes to infinity.

To construct the fault-tolerant circuit $\bar{\Phi}$ according to Theorem 1, we consider Gottesman's fault-tolerant scheme with constant overhead [6,7]. Gottesman's construction involves two different types of codes:

- (i) QLDPC codes, which store the logical qubits on which quantum computation is performed.
- (ii) Concatenated codes, which are used only for preparing certain entangled ancillary states in the QLDPC code space. These states are used for fault-tolerant logic gate implementation on the QLDPC code via gate teleportation.

We note that the overhead in ancillary state preparation using concatenated codes is not constant, i.e., it grows with code size. To make sure that the qubit overhead remains constant with respect to x , i.e., the number of qubits in Φ , a sequentialization technique is used where x logical qubits are encoded into multiple blocks of a QLDPC code and a logical gate is applied on only one (or a small number) of these blocks (see also Fig. 2).

In our generalization of Gottesman's protocol for general noise models, we assume the availability of a single-shot error correction for the QLDPC code, with constant-time complexity. However, single-shot error correction is not assumed for the concatenated codes. Fault-tolerant preparation of ancillary states using concatenated codes has been analyzed in Ref. [18]. The state preparation procedure is relatively slow, in the sense that its time complexity grows with the size of the QLDPC code. Nevertheless, this does not conflict with the use of single-shot error correction on the QLDPC code. This is because while the ancillary state is being prepared, repeated rounds of error correction are applied on the QLDPC code (see also Fig. 2). These repeated rounds of error correction ensure that logical information remains protected in the QLDPC code throughout the preparation process.

Below, we briefly describe the construction of the fault-tolerant quantum circuit $\bar{\Phi}$ according to Theorem 1.

We divide x qubits of circuit Φ into x/m blocks, each containing $m = O(\sqrt{x})$ qubits. We assume a sequential realization of Φ , where at each layer (time step), only one nontrivial gate is applied on a selected set of qubits, and the idle gate is applied on the remaining qubits. To construct $\bar{\Phi}$, we consider a QLDPC code that encodes m logical qubits in n physical qubits. We replace each block of Φ , containing m qubits, by n qubits corresponding to the QLDPC code. Then $\bar{\Phi}$ is obtained by replacing nontrivial gates in the sequential circuit Φ by the corresponding logic gate based on gate teleportation. The idle gates are replaced by multiple rounds of single-shot error correction, taking into account the time complexity of ancillary state preparation within gate teleportation. The realization of a logic gate through gate teleportation is illustrated in Fig. 2. Note that logic gates in $\bar{\Phi}$ need to be applied sequentially due to the fact that the qubit overhead in ancilla preparation is not constant but polylogarithmic (see also Fig. 2). This means that for one nontrivial logic gate, we need $O(\sqrt{x} \text{polylog}(x/\epsilon))$ ancilla qubits, where ϵ is the accuracy with which we wish to realize Φ . The sequential

implementation ensures that ancilla qubits can be reused or reinitialized, hence keeping the total qubit overhead constant. Applying parallel logic gates would mean preparing ancilla states in parallel, hence blowing up the overhead with respect to x .

The remaining sections are organized as follows. In Sec. II, we give basic definitions and review the state preparation theorem from [18]. In Sec. III, we provide a proof of Theorem 1, assuming a definition of fault-tolerant error correction. Later in Sec. IV, we use the single-shot decoding from Ref. [17] to show that the definition of fault-tolerant error correction is satisfied for QLDPC codes with constant rate and linear minimum distance, assuming general circuit noise during syndrome extraction.

II. PRELIMINARIES

In this section, we provide some basic definitions, outline the circuit model of quantum computation, and describe general circuit noise, used to derive the results in this paper. Moreover, we review the state preparation theorem from Ref. [18], which will be used later in the proof of constant overhead fault-tolerant quantum computation.

Notation. We use symbols M , L , and N for finite dimensional Hilbert spaces, and $\mathbf{L}(M, N)$ for linear operators from M to N or simply $\mathbf{L}(M)$ for linear operators on M . We use $\mathbf{D}(M)$ for quantum states on M . A qubit Hilbert space is denoted by $\mathbb{C}^2$. Quantum channels are denoted by the calligraphic letters $\mathcal{T}$, $\mathcal{P}$, $\mathcal{R}$, $\mathcal{N}$, $\mathcal{V}$, $\mathcal{W}$, etc. The identity operator is denoted by $\mathbb{1}$ and the identity quantum channel by $\mathcal{I}$.

A. Basic definitions

Let $[n] := \{1, \dots, n\}$ be a set of labels for n qubits. For any $A \subseteq [n]$, we define the space of errors with support on A as the space of linear operators acting only on qubits in A , i.e.,

$$\mathcal{E}(A) := \{(\otimes_{i \in [n] \setminus A} \mathbb{1}_i) \otimes E_A : E_A \in \mathbf{L}(M^{\otimes |A|})\} \subseteq \mathbf{L}(M^{\otimes n}). \quad (1)$$

Using $\mathcal{E}(A)$, we define the space of operators with *weight* $z \leq n$ qubits as follows.

Definition 1 (Weight of an operator). The space of operators with weight $z \leq n$ is defined as

$$\mathcal{E}(n, z) := \text{span}\{E : E \in \mathcal{E}(A) \text{ for some } A \subseteq [n] \text{ s.t. } |A| \leq z\}. \quad (2)$$

Definition 2 (Weight of a superoperator). The space of superoperators with weight $z \leq n$ is defined as

$$\mathcal{E}(n, z)(\cdot)\mathcal{E}(n, z)^* = \text{span}\{E_1(\cdot)E_2 : E_1 \in \mathcal{E}(n, z), E_2 \in \mathcal{E}(n, z)^*\}. \quad (3)$$

We note that in Definition 2, the operators E_1 and E_2 both have weight z but they may be supported on completely different subsets of the n qubits. We have adopted this broader definition so as not to restrict error maps to quantum channels, but rather consider them as general superoperators. This broader definition simplifies the analysis of fault-tolerant error correction under general noise models.

We will use the diamond norm as a measure of distance between two superoperators [5, 19, 20].

Definition 3 (Diamond norm). For any superoperator $\mathcal{T} : \mathbf{L}(M) \rightarrow \mathbf{L}(N)$, the diamond norm $\|\mathcal{T}\|_\diamond$ is defined as

$$\|\mathcal{T}\|_\diamond := \sup_G \|\mathcal{I}_G \otimes \mathcal{T}\|_1, \quad (4)$$

where G is a reference system and the 1-norm is given by

$$\|\mathcal{T}\|_1 := \sup_{\substack{\|\rho\|_1 \leq 1 \\ \rho \in \mathbf{L}(M)}} \|\mathcal{T}(\rho)\|_1. \quad (5)$$

B. Quantum circuits and noise model

We consider joint quantum computation on classical and quantum bits. Any quantum computation is performed using a quantum circuit, which is constructed from a gate set acting on classical and quantum bits. We consider a gate set $\mathbf{A}$, containing a universal set of gates on classical bits, and a set of quantum gates, including state preparation, quantum measurement, a universal set of unitary gates, and the corresponding classically controlled unitary gates.

As in Ref. [18, Sec. 2], we consider a hybrid model of a quantum gate, acting on classical and quantum bits. We divide classical bits as control bits and operational bits. Let l be the number of control bits and let o and o' be the number of operational bits at the input and output, respectively. Let q and q' be the number of qubits at the input and output, respectively. For $\mathbf{u} \in \{0, 1\}^l$, define a classical linear map $\mathcal{B}^{\mathbf{u}} : \mathbb{B}^{\otimes l} \rightarrow \mathbb{B}^{\otimes l}$, $\mathcal{B}^{\mathbf{u}}(\cdot) := |\mathbf{u}\rangle\langle\mathbf{u}|(\cdot)|\mathbf{u}\rangle\langle\mathbf{u}|$. Then, the gate g realizes the quantum channel

$$\mathcal{T}_g := \sum_{\mathbf{u} \in \{0, 1\}^l} \mathcal{B}^{\mathbf{u}} \otimes \mathcal{T}^{\mathbf{u}}, \quad (6)$$

where $\mathcal{T}^{\mathbf{u}}$ is a quantum channel from $\mathbb{B}^{\otimes o} \otimes \mathbf{L}((\mathbb{C}^2)^{\otimes q})$ to $\mathbb{B}^{\otimes o'} \otimes \mathbf{L}((\mathbb{C}^2)^{\otimes q'})$. Here, $\mathcal{T}_g$ signifies transformation associated with the gate g .

We consider a definition of quantum circuit based on layered implementation of gates, as given in Ref. [5] (see also Ref. [18, Sec. 2]).

Definition 4 (Quantum circuit). A quantum circuit Φ of depth d is a collection of the following objects (see also Fig. 3):

- (1) A sequence of finite sets $\Delta_0, \dots, \Delta_d$, called *layers*. The zeroth layer Δ_0 is called the *input* of the circuit, and the final layer Δ_d is the *output*. Each layer Δ_i (for $i = 0, \dots, d$) decomposes as $\Delta_i = \Delta_i^c \cup \Delta_i^q$, where Δ_i^c contains *classical wires* and Δ_i^q contains *quantum wires*. If the circuit has only classical input, then $\Delta_0^q = \emptyset$; if the circuit has no input, then $\Delta_0 = \emptyset$.
- (2) Partitions of each set Δ_{i-1} , $i = 1, \dots, d$ into ordered subsets (registers) $A_{i1}, \dots, A_{is_i}$, and a corresponding partition of a superset $\Delta'_i \supseteq \Delta_i$ into registers $A'_{i1}, \dots, A'_{is_i}$. Each register A_{ij}, A'_{ij} contains l_{ij} control bits, q_{ij}, q'_{ij} qubits, and o_{ij}, o'_{ij} operational bits, respectively.
- (3) A collection of gates $g_{i,j}$ drawn from the gate set $\mathbf{A}$, each realizing a quantum channel

$$\mathcal{T}_{g_{i,j}} : \mathbb{B}^{\otimes l_{ij}} \otimes \mathbb{B}^{\otimes o_{ij}} \otimes \mathbf{L}((\mathbb{C}^2)^{\otimes q_{ij}}) \rightarrow \mathbb{B}^{\otimes l_{ij}} \otimes \mathbb{B}^{\otimes o'_{ij}} \otimes \mathbf{L}((\mathbb{C}^2)^{\otimes q'_{ij}}).$$

The gate $g_{i,j}$ or their numbers (i,j) are called the *locations* of the quantum circuit. Note that for a fixed i , $1 \leq i \leq d$, the set of locations $\{g_{i,j} : 1 \leq j \leq s_i\}$ corresponds to the locations in a single layer of the quantum circuit.

The *size* of Φ , denoted by $|\Phi|$, is defined as the total number of *locations* it contains.

A quantum circuit Φ according to Definition 4 realizes a quantum channel

$$\mathcal{T} = \mathcal{T}_d \circ \dots \circ \mathcal{T}_1, \quad (7)$$

where $\mathcal{T}_i$ acts as follows on any $\rho \in \mathbf{L}(\mathbb{C}^{2^{\otimes |\Delta_{i-1}|}})$,

$$\mathcal{T}_i(\rho) = \text{Tr}_{\Delta'_i \setminus \Delta_i} \left[\mathcal{T}_{g_{i,1}}[A'_{i1}; A_{i1}] \otimes \dots \otimes \mathcal{T}_{g_{i,s_i}}[A'_{is_i}; A_{is_i}] \right] (\rho), \quad (8)$$

where $\mathcal{T}_{g_{i,j}}[A'_{ij}; A_{ij}]$ denotes the quantum channel $\mathcal{T}_{g_{i,j}}$, applied to the register A_{ij} , the output of which is stored in the register A'_{ij} .

Remark 1. Note that in Definition 4, all the qubits in Δ_{i-1} , $i = 1, \dots, d$, are acted upon by a gate at the i th layer. We suppose that the idle gate acts on each waiting qubit. This is natural for capturing the effect of noise on the waiting qubits (see Sec. II C).

C. General circuit noise

Let $\delta \in [0, 1]$ be an error rate. When general circuit noise with an error rate δ acts on a quantum circuit Φ , we have that:

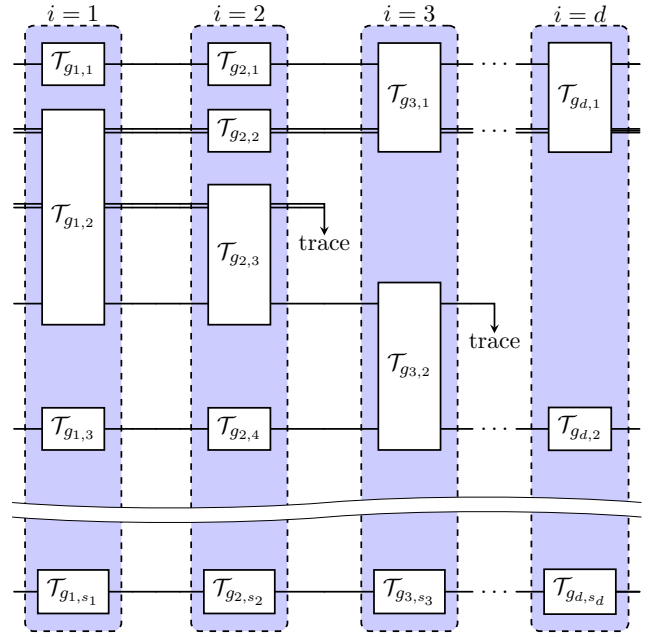


FIG. 3. An example of quantum circuit according to Definition 4.

- (1) Purely classical elements of Φ , i.e., the gates with only classical input-output, are implemented perfectly without any error.
- (2) A gate $g \in \mathbf{A}$ with the hybrid classical-quantum input as in Eq. (6) is replaced by an arbitrary $\tilde{\mathcal{T}}$ such that

$$\tilde{\mathcal{T}}_g = \sum_{\mathbf{u} \in \{0,1\}^l} \mathcal{B}^{\mathbf{u}} \otimes \tilde{\mathcal{T}}^{\mathbf{u}}, \quad (9)$$

where for all $\mathbf{u} \in \{0,1\}^l$, $\tilde{\mathcal{T}}^{\mathbf{u}}$ is an arbitrary quantum channel with the same input and output space as $\mathcal{T}^{\mathbf{u}}$ (in particular, classical registers remain classical), and $\|\tilde{\mathcal{T}}^{\mathbf{u}} - \mathcal{T}^{\mathbf{u}}\|_{\diamond} \leq \delta$. Note that a noisy idle gate realizes a channel $\tilde{\mathcal{T}}_g$, such that $\|\tilde{\mathcal{T}}_g - \mathcal{I}\|_{\diamond} \leq \delta$.

A quantum circuit Φ with the noise rate δ realizes a quantum channel in the set,

$$\begin{aligned} \text{Trans}(\Phi, \delta) &:= \left\{ \tilde{\mathcal{T}}_d \circ \dots \circ \tilde{\mathcal{T}}_1 : \tilde{\mathcal{T}}_i = \text{Tr}_{\Delta'_i \setminus \Delta_i} \circ \right. \\ &\quad \left[\tilde{\mathcal{T}}_{g_{i,1}}[A'_{i1}; A_{i1}] \otimes \dots \otimes \tilde{\mathcal{T}}_{g_{i,s_i}}[A'_{is_i}; A_{is_i}] \right], \\ &\quad \text{with } \tilde{\mathcal{T}}_{g_{i,j}} \text{ being a } \delta\text{-noisy version of } \mathcal{T}_{g_{i,j}} \\ &\quad \left. \text{as in Eq. (9)}. \right\}. \end{aligned} \quad (10)$$

Here, “Trans” in $\text{Trans}(\Phi, \delta)$ stands for “Transformation.”

D. Fault-tolerant state preparation

One of the ingredients for the proof of our main result in Theorem 1 is a fault-tolerant state preparation method for preparing code states of QLDPC codes. To do this, we use a state preparation method given in Ref. [18], which can prepare any n -qubit state up to an error of weight $O(n\delta)$, under general circuit noise with error rate δ .

Before stating the state preparation theorem, we first recall the definition of an adversarial channel with noise parameter δ from Ref. [18]. We also present, for later reference, a result from Ref. [18, Theorem 67] that states that any adversarial channel with parameter δ can be approximated by a superoperator of weight $O(n\delta)$.

Definition 5 (Adversarial channels). A quantum channel $\mathcal{V} : \mathbf{L}(M^{\otimes n}) \rightarrow \mathbf{L}(M^{\otimes n})$ is said to be an adversarial channel with parameter δ if for any $t > 0$, there exists a superoperator $\mathcal{V}_t$ of weight t , i.e., $\mathcal{V}_t \in \mathcal{E}(n, t)(\cdot)\mathcal{E}(n, t)^*$ (see Def. 2), such that

$$\|\mathcal{V} - \mathcal{V}_t\|_\diamond \leq \sum_{j>t} \binom{n}{j} \delta^j. \quad (11)$$

Lemma 1. For any adversarial channel $\mathcal{V} : \mathbf{L}(M^{\otimes n}) \rightarrow \mathbf{L}(M^{\otimes n})$ with parameter δ , there exists a superoperator $\mathcal{V}'$ of weight $5n\delta$, such that

$$\|\mathcal{V} - \mathcal{V}'\|_\diamond \leq \exp\left(-\frac{n\delta}{3}\right). \quad (12)$$

Proof. From the definition of adversarial channel, there exists a superoperator $\mathcal{V}_t$ of weight t , such that

$$\|\mathcal{V} - \mathcal{V}_t\|_\diamond \leq \sum_{j>t} \binom{n}{j} \delta^j \quad (13)$$

$$= (1 + \delta)^n \sum_{j>t} \binom{n}{j} \left(\frac{\delta}{1 + \delta}\right)^j \left(\frac{1}{1 + \delta}\right)^{n-j}. \quad (14)$$

We now consider n independent and identically distributed (IID) random variables $X_1, X_2, \dots, X_n$, such that $X_i \in \{0, 1\}$, and $\Pr(X_i = 1) = \delta/(1 + \delta)$, for all $i \in [n]$. We define $X = \sum_i X_i$; then

$$\Pr(X > t) = \sum_{j>t} \binom{n}{j} \left(\frac{\delta}{1 + \delta}\right)^j \left(\frac{1}{1 + \delta}\right)^{n-j}. \quad (15)$$

Therefore, from Eqs. (16) and (15), and $(1 + \delta)^n \leq e^{n\delta}$, we obtain

$$\|\mathcal{V} - \mathcal{V}_t\|_\diamond \leq e^{n\delta} \Pr(X > t). \quad (16)$$

For a $\beta > 0$, we take $t = (1 + \beta)n\delta/(1 + \delta)$. Then, using the multiplicative Chernoff bound, we obtain

$$\Pr(X > t) \leq \exp\left(-\frac{\beta^2}{2 + \beta} \frac{n\delta}{1 + \delta}\right) \leq \exp\left(-\frac{\beta^2 n\delta}{2(2 + \beta)}\right). \quad (17)$$

By taking $\beta = 4$ (i.e., $t \leq 5n\delta$), from Eqs. (16) and (17), we have that Eq. (12) is satisfied for $\mathcal{V}' = \mathcal{V}_t$. ■

Below, we present the state preparation theorem from Ref. [18, Theorem 1].

Theorem 2 (State preparation). Consider a quantum circuit Φ with classical input and n -qubit output. Let δ be a fixed noise rate, considering general circuit noise. Then, for any positive integer k , there exists a quantum circuit $\bar{\Phi}$, with the same input and output systems as Φ and of size $|\Phi| \cdot \text{poly}(k)$, such that for any noisy version $\tilde{\mathcal{T}}_{\bar{\Phi}} \in \text{Trans}(\bar{\Phi}, \delta)$,

$$\|\tilde{\mathcal{T}}_{\bar{\Phi}} - \mathcal{W} \circ \Phi\|_\diamond \leq O(|\Phi| \sqrt{(c\delta)^k}), \quad (18)$$

where $\mathcal{W}$ is an adversarial channel with parameter $O(\delta)$, and c is a universal constant.

Remark 2. The fault-tolerant circuit $\bar{\Phi}$ corresponding to the state preparation circuit Φ in Theorem 2 is constructed using a fault-tolerant scheme [5, 18], which consists of the following three points:

- (a) *Quantum codes.* A sequence of quantum codes $D_k, k = 1, 2, \dots$, encoding one logical qubit in n_k physical qubits, with D_1 being the trivial code, i.e., the corresponding $n_k = 1$.
- (b) *Logic gate.* For any gate g from a gate set, a sequence of quantum circuits $\Psi_{g,k}, k = 1, 2, \dots$, such that the corresponding noisy version, with noise parameter $\delta > 0$, represents (simulates) the gate g in codes D_k , with an accuracy $(c\delta)^k$, for a constant $c > 0$.
- (c) *Interface.* A sequence of circuits $\Gamma_{k,l}, k, l = 1, 2, \dots$, such that the corresponding noisy version with noise parameter δ , represents the identity gate in codes D_k and D_l , with accuracy $(c\delta)^{\min\{k,l\}}$.

Moreover, it is required that the size of the quantum circuits $\Psi_{g,k}$ and $\Gamma_{k,l}$ scale polynomially, with respect to k and l . It has been shown that a fault-tolerant scheme can be built using concatenated codes $\mathcal{C}_r, r = 0, 1, 2, \dots$, where r is the level of concatenation, which is related to the parameter k in point (a) of the fault-tolerant scheme as $r \sim \log k$ [18]. For $r = 0$, $\mathcal{C}_r$ is the trivial code and $\mathcal{C}_1$ is the base code for concatenation. It is required that $\mathcal{C}_1$ has minimum distance at least 5, i.e., it corrects at least two errors.

The interface circuit $\Gamma_{k,l}$, corresponding to $l = 1$, plays an important role in the construction of the fault-tolerant circuit $\bar{\Phi}$ from Theorem 2. Note that $\Gamma_{k,1}$ maps a logical qubit encoded in n_k physical qubits to a bare logical qubit, while preserving the logical information with a reasonable accuracy $O(\delta)$. The fault-tolerant circuit $\bar{\Phi}$ corresponding to Φ and integer k in Theorem 2 can be taken as

$$\bar{\Phi} := \otimes_{i=1}^n \Gamma_{k,1} \circ \Phi_k, \quad (19)$$

where Φ_k is obtained from Φ by replacing each gate g with the corresponding fault-tolerant logic gate $\Psi_{g,k}$. As both $\Psi_{g,k}$ and $\Gamma_{k,l}$ have size scaling polynomially in k and l , respectively, it follows that the size of $\bar{\Phi}$ is $|\Phi| \cdot \text{poly}(k)$, and its depth is $d \cdot \text{poly}(k)$, where d is the depth of Φ .

III. CONSTANT-OVERHEAD FAULT-TOLERANT QUANTUM COMPUTING FOR GENERAL CIRCUIT NOISE

In this section, we prove our main result, stated in Theorem 1. We will first formally state our constant-overhead result in Theorem 3, and then provide a proof.

Theorem 3. Let Φ be a quantum circuit, with classical input and output, working on x qubits and $|\Phi| = \text{poly}(x)$. For any $\epsilon > 0$, and asymptotic overhead $\alpha > 1$, there exists a constant error rate $\delta_{\text{th}} > 0$, and a quantum circuit Φ_ϵ , $|\Phi_\epsilon| = \text{poly}(x)$, with the same input and output systems as Φ , and working on $x' = \alpha x + o(x)$ qubits, such that for any $\tilde{\mathcal{T}}_{\Phi_\epsilon} \in \text{Trans}(\Phi_\epsilon, \delta)$, $\delta < \delta_{\text{th}}$,

$$\|\tilde{\mathcal{T}}_{\Phi_\epsilon} - \mathcal{T}_\Phi\|_1 \leq \epsilon. \quad (20)$$

A. Proof of Theorem 3

We consider the fault-tolerant construction from Refs. [6,7], which requires Φ to be sequential in the following sense: at any layer of the circuit Φ , only one nontrivial gate is applied on a selected set of qubits, and the idle gate is applied on the remaining qubits (see also Remark 1). This means that we need to map Φ to a sequential circuit Φ_{seq} . Any circuit Φ can be mapped to a sequential circuit Φ_{seq} by incurring only a linear overhead in the size of the circuit, $|\Phi_{\text{seq}}| = O(x)|\Phi|$. Note that the linear overhead is due to the additional idle gates, which are accounted for in our circuit model. Therefore, without loss of generality, we suppose Φ to be a sequential circuit.

We consider a family of QLDPC codes with rate $R = 2/(1 + \alpha)$, where $\alpha > 1$ is the desired asymptotic overhead and take a code C of type (n, m) in this family (note that asymptotically $n = m/R$). For this family of QLDPC codes, we suppose fault-tolerant realizations of error correction and logic gates, as described in the following two paragraphs, respectively.

1. Fault-tolerant error correction

Consider h code blocks of the code C and a joint state $|\psi\rangle$ in the code space of these h blocks. Let $\mathcal{V}$ be an error map on the state $|\psi\rangle\langle\psi|$, and suppose that it has stabilizer reduced weight sn for a constant $0 < s < 1$, with respect to each code block (see Definitions 7–9 below for definition of the stabilizer reduced weight). Then, when the error correction circuit with general noise is applied on every code block, it maps $\mathcal{V}(|\psi\rangle\langle\psi|)$ to $\mathcal{V}'(|\psi\rangle\langle\psi|)$ with high accuracy, where the stabilizer reduced weight of $\mathcal{V}'$ with respect to each code block is given by $s'n$, with $s' < s$. More precisely, let Φ_{EC}^i be the instance of the error correction circuit applied on the block indexed $i \in [h]$, and let $\tilde{\mathcal{T}}_{\Phi_{\text{EC}}}^i \in \text{Trans}(\Phi_{\text{EC}}^i, \delta)$, $i \in [h]$ be a noisy realization under general circuit noise. Then, we have

$$\begin{aligned} & \|(\otimes_{i \in [h]} \tilde{\mathcal{T}}_{\Phi_{\text{EC}}}^i) \circ \mathcal{V}(|\psi\rangle\langle\psi|) - \mathcal{V}'(|\psi\rangle\langle\psi|)\|_1 \\ & \leq O(h \exp(-O(n\delta))). \end{aligned} \quad (21)$$

We provide in Theorem 5 a fault-tolerant error correction circuit Φ_{EC} , satisfying Eq. (21). To this end, we show that the single-shot error correction for quantum Tanner codes from Ref. [17] satisfies Eq. (21), while considering general circuit noise for syndrome extraction. We note that single-shot error correction requires only one round of noisy syndrome measurement, which can be done by a circuit of constant depth (independent of n) for QLDPC codes. Moreover, the classical decoding algorithm, estimating the error, is also of constant depth [17]. Thus, single-shot error correction is executed in constant time.

2. Realization of logic gates

The logic gate corresponding to preparing the $|0\rangle$ state is realized using our state preparation procedure given in Theorem 2. For a logic unitary gate, we use gate teleportation as in Refs. [6,7]. To realize the logical version of a unitary gate U_g between logical qubits in the same code block, one needs to prepare as ancilla the entangled code state

$$|\psi'_{U_g}\rangle = V \otimes V |\psi_{U_g}\rangle, \quad (22)$$

where V is the encoding isometry corresponding to the QLDPC code, and $|\psi_{U_g}\rangle$ is a $2m$ -qubit entangled state,

$$|\psi_{U_g}\rangle = (\mathbb{1} \otimes U_g) \left(\frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} \right)^{\otimes m}, \quad (23)$$

with U_g acting on a selected set of qubits in the second group of m qubits and identity is applied everywhere else. A similar state is needed when U_g acts on qubits in different blocks. We use Theorem 2 to prepare the code state $|\psi'_{U_g}\rangle$ fault tolerantly.

During the gate teleportation, one performs logical Bell measurements between the code block, and the first code block of the ancilla state. After the Bell measurement, the state of the remaining ancilla block is given by $U_{\text{cor}}^\dagger U_g^L |\psi\rangle$, where $|\psi\rangle$ is the logical state corresponding to the code block prior to gate teleportation, U_g^L is the logical unitary corresponding to U_g and U_{cor} is a correction unitary. If U_g is in the Clifford group, then the correction is a Pauli gate which is transversal [21, Ch. 6]. If U_g is in the second level of the Clifford hierarchy (e.g., U_g is the T gate), then the correction is a Clifford gate that we can apply in a transversal way using another state preparation for this correction unitary.

We note that the logical Bell measurement can be performed by applying a transversal CNOT gate between the code blocks, then a transversal Hadamard gate on the first block, and then a transversal single qubit Pauli measurements on the two code blocks [21, Ch. 6]. For a detailed discussion of the implementation of logic gates via ancillary preparation and gate teleportation, we refer to Ref. [21, Ch. 6].

3. Layout of the circuit Φ_ϵ

The quantum circuit Φ_ϵ simulating Φ is constructed as follows. We divide x qubits of Φ into x/m blocks, each containing $m = O(\sqrt{x})$ qubits [22]. We replace each block of m qubits by n qubits corresponding to a code block of the QLDPC code C . The quantum circuit Φ_ϵ is obtained by replacing each gate in the circuit Φ by the corresponding logic gate, which acts within one block or between a pair of blocks, each containing n qubits.

As Φ is a sequential circuit, a layer in Φ realizes only one nontrivial gate. In Φ_ϵ , the logical version of a layer in Φ is realized in several layers at the physical level. The depth of the logical layer is determined by the depth of the nontrivial logic gate applied at that layer. While the nontrivial logic gate is applied on the respective blocks in Φ_ϵ , repeated error correction is applied on the remaining blocks. We illustrate in Fig. 2 the implementation of a logic unitary gate, using gate teleportation, when U_g is applied on a set of logical qubits within one code block.

The logic gate corresponding to a unitary U_g is realized in the following two steps (see also Fig. 2):

- (i) *Ancilla preparation.* During this step, the required ancilla code state $|\psi'_{U_g}\rangle$ from Eq. (22) is prepared. While the ancilla state is prepared, one keeps applying error correction on the code blocks of the QLDPC codes (see Fig. 2). We know that there exists a circuit Ψ'_{U_g} , working on $O(n)$ qubits, which prepares $|\psi'_{U_g}\rangle$, with size $|\Psi'_{U_g}| = O(\text{poly}(n))$ and depth $d' = O(\text{poly}(n))$ [23, Ch. 4] [24]. To prepare $|\psi'_{U_g}\rangle$, we consider the fault-tolerant circuit $\bar{\Psi}'_{U_g}$ associated with Ψ'_{U_g} , according to Theorem 2. Let

ρ be the output of $\bar{\Psi}'_{U_g}$, with general circuit noise. Then, from Eq. (18), we have

$$\|\rho - \mathcal{V}(|\psi'_{U_g}\rangle\langle\psi'_{U_g}|)\|_1 \leq O(|\Psi'_{U_g}|)(c\delta)^{k/2}, \quad (24)$$

where $\mathcal{V}$ is a superoperator of weight $O(n\delta)$ and k is the parameter from Theorem 2. Therefore, $\bar{\Psi}'_{U_g}$ prepares the code state $|\psi'_{U_g}\rangle$ up to an error of weight $O(n\delta)$, with very high accuracy given that k is sufficiently large. From Theorem 2 and Remark 2, note that the size and the depth of $\bar{\Psi}'_{U_g}$ are given by

$$|\bar{\Psi}'_{U_g}| = |\Psi'_{U_g}| \text{poly}(k) = O(\text{poly}(n) \text{poly}(k)) \quad (25)$$

$$\bar{d}' = d' O(\text{poly}(k)) = O(\text{poly}(n) \text{poly}(k)) \quad (26)$$

- (ii) *Gate teleportation.* The prepared ancilla state is then used to implement the logic gate on the selected code block, using gate teleportation. On the remaining blocks, error correction circuits are applied.

From Eq. (26), we note that the time complexity of the ancillary state preparation increases with n , in contrast with the single-shot error correction on the QLDPC code. This, however, does not pose a problem since while the ancillary state is being prepared, repeated rounds of error correction are performed on the QLDPC code (see Fig. 2); thereby logical information remains protected throughout the state preparation.

4. Error bound on computation

Our state preparation procedure together with the transversality of the gates used in gate teleportation, imply that only an error of weight $c_g \delta n$, for a constant $c_g > 0$, is introduced during the implementation of the logic gate. By choosing the parameter δ small enough so that $s' + c_g \delta < s$, we can make sure that error can be suppressed in the next round of error correction, as given in Eq. (21). Therefore, after the t th logical layer of the quantum circuit Φ_ϵ , with general noise (supposing that the last layer is error correction), we obtain a quantum state ρ_t , which is close to the desired code state $|\psi_t\rangle\langle\psi_t|$ (i.e., the state at the t th step of Φ_ϵ without noise) in the following sense:

$$\begin{aligned} \|\rho_t - \mathcal{V}'(|\psi_t\rangle\langle\psi_t|)\|_1 &\leq O(\bar{d}' \sqrt{x} \exp(-O(\sqrt{x}\delta))) \\ &\quad + O(t|\Psi_{U_g}|)(c\delta)^{k/2}, \end{aligned} \quad (27)$$

where $\mathcal{V}'$ is a superoperator of weight $s'n$, with respect to each code block. The first error term on the right-hand side of Eq. (27) is due to fault-tolerant error correction (see Eq. (21)), and the second error term is due to the ancilla

state preparation [see Eq. (24)]. For the second term, using $|\Psi_{U_g}\rangle = O(\text{poly}(n)) = O(\text{poly}(\sqrt{x}))$, and $t = \text{poly}(x)$, it suffices to choose $k = \text{polylog}(x/\epsilon)$ to make this term of order ϵ . Therefore, from Eqs. (25) and (26), the size and depth of the state preparation circuit are given by

$$|\overline{\Psi'}_{U_g}| = O\left(\text{poly}(x) \text{polylog}\left(\frac{1}{\epsilon}\right)\right), \quad (28)$$

$$\overline{d'} = O\left(\text{poly}(x) \text{polylog}\left(\frac{1}{\epsilon}\right)\right). \quad (29)$$

Finally, since $t = \text{poly}(x)$, the first term approaches to zero as $x \rightarrow \infty$.

5. Proof of constant qubit overhead

We now count the number of qubits in the quantum circuit Φ_ϵ :

- (i) *Data qubits.* The number of data qubits corresponding to all the code blocks is given by x/R .
- (ii) *Ancilla qubits for generator measurements.* The number of ancilla qubits used for error correction in each block is $n - m$. Therefore, the total number of ancilla qubits needed for error correction is given by $(n - m)x/m = x((1/R) - 1)$.
- (iii) *Ancilla qubits for state preparation.* As discussed before, we can take $k = \text{polylog}(x/\epsilon)$ to achieve an error of order ϵ in ancilla state preparation. Since the qubit overhead of preparation according to Theorem 2 is $\text{poly}(k)$, the number of qubits needed to prepare the ancilla state is given by

$$O\left(n \text{polylog}\left(\frac{x}{\epsilon}\right)\right) = O\left(\sqrt{x} \text{polylog}\left(\frac{x}{\epsilon}\right)\right). \quad (30)$$

Therefore, the total number of qubits is given by

$$x' = 2x \frac{1}{R} - x + O\left(\sqrt{x} \text{polylog}\left(\frac{x}{\epsilon}\right)\right) = \alpha x + o(x). \quad (31)$$

IV. SINGLE-SHOT ERROR CORRECTION FOR QLDPC CODES AGAINST GENERAL ERRORS

For fault-tolerant error correction, one needs to consider the errors on both data qubits and syndrome bits as errors also happen during the syndrome extraction procedure. Usually, syndrome errors are dealt with by extracting several rounds of syndromes [25,26]. Single-shot error correction uses only a single round of noisy syndrome extraction to perform error correction [27,28]. Recently, it has been shown that the (asymptotically) good QLDPC codes of Ref. [15], namely, quantum Tanner codes, support a single-shot decoding [17], where Pauli errors are applied

on the data qubits, and bit-flip errors on the syndrome bits. In this section, we generalize the single-shot error correction for general linear error operators on data qubits. This generalization will be needed for the fault-tolerant error correction under general circuit noise, given in Theorem 5 in the next section.

Below, we provide a definition of QLDPC codes.

Definition 6 ((r,s)-QLDPC codes [6]). A stabilizer code is said to be an (r,s) -QLDPC code if there exists a choice of stabilizer generators $\{M_1, \dots, M_l\}$ such that the weight of M_i is at most r , for all $i \in [l]$, and each qubit is nontrivially involved in at most s generator measurements.

In what follows, we fix a Calderbank-Shor-Steane (CSS) QLDPC code C of type (n,m) , i.e., encoding m logical qubits in n physical qubits, and with minimum distance $d_{\min}$. Moreover, we suppose that C admits a single-shot decoding procedure. For example, one can choose a quantum Tanner code from Ref. [15] and the decoder analysis of Ref. [17]. We denote the generating set $\{M_1, \dots, M_l\}$, $l < n$, as in Definition 6 (generally, $l = n - m$). The encoding channel of C is given by $V(\cdot)V^\dagger$, where $V: (\mathbb{C}^2)^{\otimes m} \rightarrow (\mathbb{C}^2)^{\otimes n}$ is an isometry, and the code space $\overline{L} := \text{Im}(V) \subseteq (\mathbb{C}^2)^{\otimes n}$.

The error correction map is given by a classical controlled unitary

$$\mathcal{U}_{\text{ec}}: \mathbb{B}^{\otimes l} \otimes \mathbf{L}((\mathbb{C}^2)^{\otimes n}) \rightarrow \mathbb{B}^{\otimes l} \otimes \mathbf{L}((\mathbb{C}^2)^{\otimes n}),$$

where $\mathbb{B}^{\otimes l}$ refers to the syndrome bits, i.e., the outcome of generator measurements $\{M_1, \dots, M_l\}$. For stabilizer codes, $\mathcal{U}_{\text{ec}}$ is a controlled Pauli, where l bits act as control and n qubits as target. For simplicity of notation, we define $\mathcal{U}_{\text{ec}} := U_{\text{ec}}(\cdot)U_{\text{ec}}^\dagger$, where U_{ec} is a controlled Pauli, with l qubit control and n -qubit target.

In the following definition, we recall the stabilizer reduced error weight.

Definition 7 (Stabilizer reduced error weight [17]). Consider a CSS code of type (n,m) , and let $X^{\mathbf{e}_x}, \mathbf{e}_x \in \{0,1\}^n$ be a Pauli- X error. Let $|\mathbf{e}_x|$ denote the Hamming weight of $\mathbf{e}_x$. The stabilizer reduced weight $|\mathbf{e}_x|_{\text{red}}$ of $\mathbf{e}_x$ is defined as

$$|\mathbf{e}_x|_{\text{red}} := \min\{|\mathbf{e}'_x| : X^{\mathbf{e}'_x} \text{ is equivalent to } X^{\mathbf{e}_x} \text{ up to stabilizer multiplication}\}.$$

Similarly, we can define the stabilizer reduced weight for Pauli- Z errors. The stabilizer reduced weight of the total error $\mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z)$ is defined as $|\mathbf{e}|_{\text{red}} := \max\{|\mathbf{e}_x|_{\text{red}}, |\mathbf{e}_z|_{\text{red}}\}$.

Definitions 8 and 9 extend the definition of the reduced error weight to linear error operators and superoperators, respectively. To this end, we consider error operators that

act jointly on the n physical qubits of the code and on a $\bar{n}$ -qubit auxiliary system. This is done keeping in mind that we will consider error correction of multiple code blocks in parallel; therefore, we would need a notion of reduced error weight with respect to each code block.

Definition 8. Consider a CSS code of type (n, m) . Let $E \in \mathbf{L}((\mathbb{C}^2)^{\otimes \bar{n}} \otimes (\mathbb{C}^2)^{\otimes n})$ be an operator, jointly acting on a code block containing n physical qubits and on an auxiliary system containing $\bar{n}$ qubits. We expand E as

$$E = \sum_{\mathbf{e}_x, \mathbf{e}_z \in \{0,1\}^n} E_{\mathbf{e}_x, \mathbf{e}_z} \otimes X^{\mathbf{e}_x} Z^{\mathbf{e}_z}, \quad (32)$$

where $E_{\mathbf{e}_x, \mathbf{e}_z} := 1/2^n \text{Tr}_{[n]}(E(\mathbb{1}_{\bar{n}} \otimes Z^{\mathbf{e}_z} X^{\mathbf{e}_x})) \in \mathbf{L}((\mathbb{C}^2)^{\otimes \bar{n}})$, where $\text{Tr}_{[n]}$ refers to tracing out the n qubits in the code block. The reduced weight of E with respect to the code block is defined as

$$|E|_{\text{red}} := \max\{|\mathbf{e}|_{\text{red}}, \mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z) \in \{0,1\}^{2n} : E_{\mathbf{e}_x, \mathbf{e}_z} \neq 0\}.$$

Definition 9. For a superoperator $\mathcal{V} : \mathbf{L}((\mathbb{C}^2)^{\otimes \bar{n}} \otimes (\mathbb{C}^2)^{\otimes n}) \rightarrow \mathbf{L}((\mathbb{C}^2)^{\otimes \bar{n}} \otimes (\mathbb{C}^2)^{\otimes n})$, we say $|\mathcal{V}|_{\text{red}} \leq t$ if there exists a decomposition

$$\mathcal{V} = \sum_i E_i(\cdot) E_i^\dagger, \quad (33)$$

such that $\max_i\{|E_i|_{\text{red}}, |E_i'|_{\text{red}}\} \leq t$.

A. Results for single-shot error correction

Let $\Pi_\sigma \in \mathbf{L}((\mathbb{C}^2)^{\otimes n})$ be the projector on the space corresponding to the syndrome $\sigma := (\sigma_1, \dots, \sigma_l) \in \{0,1\}^l$, i.e.,

$$\Pi_\sigma = \prod_{i \in [l]} \frac{1}{2} (1 + (-1)^{\sigma_i} M_i). \quad (34)$$

Let $\sigma_{\mathbf{e}} = (\sigma_{\mathbf{e}_x}, \sigma_{\mathbf{e}_z})$ be the syndrome corresponding to the Pauli error given by $\mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z)$, i.e., $\sigma_{\mathbf{e}_x} = H_Z \mathbf{e}_x$, $\sigma_{\mathbf{e}_z} =$

$H_X \mathbf{e}_z$, where H_Z and H_X are Z - and X -type parity-check matrices of the CSS code C , respectively. We have for any $|\psi\rangle \in \bar{L}$,

$$\Pi_\sigma X^{\mathbf{e}_x} Z^{\mathbf{e}_z} |\psi\rangle = \alpha_{\sigma, \sigma_{\mathbf{e}}} X^{\mathbf{e}_x} Z^{\mathbf{e}_z} |\psi\rangle, \quad (35)$$

where $\alpha_{\sigma, \sigma_{\mathbf{e}}} = 1$ if $\sigma = \sigma_{\mathbf{e}}$ and $\alpha_{\sigma, \sigma_{\mathbf{e}}} = 0$ if $\sigma \neq \sigma_{\mathbf{e}}$. The quantum channel corresponding to the (perfect) syndrome measurement is given by

$$\mathcal{M}_{\text{syn}}(\cdot) := \sum_{\sigma \in \{0,1\}^l} |\sigma\rangle\langle\sigma| \otimes \Pi_\sigma(\cdot) \Pi_\sigma. \quad (36)$$

The following channel captures the effect of the syndrome error $\mathbf{e}^{\text{syn}} \in \{0,1\}^l$, where the recorded syndrome is σ , but the realized projective measurement is $\Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}}$:

$$\mathcal{M}_{\text{syn}}^{\mathbf{e}^{\text{syn}}}(\cdot) := \sum_{\sigma \in \{0,1\}^l} |\sigma\rangle\langle\sigma| \otimes \Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}}(\cdot) \Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}}. \quad (37)$$

Below, we present the one shot-decoding theorem from Ref. [17, Theorem 3.5], stated in a slightly different language. We have taken constant $\alpha = 0$ in Ref. [17, Theorem 3.5]. Informally, the theorem states that if the data error and the syndrome errors are bounded, then the error correction procedure suppresses the input error, even though it may not completely remove it (see Fig. 4).

Theorem 4 (Single-shot decoding [17]). There exist constants $R > 0$, $S > 0$, and $\beta > 0$ with $\beta S n + R n < d_{\min}/2$, such that the following holds.

Let $|\psi\rangle \in \bar{L}$ be a quantum state in the code space of the QLDPC code C . Suppose that the data error is $\mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z) \in \{0,1\}^{2n}$ with reduced weight $|\mathbf{e}|_{\text{red}} \leq \beta S n + R n$ and the syndrome error is $\mathbf{e}^{\text{syn}} \in \{0,1\}^l$ with weight

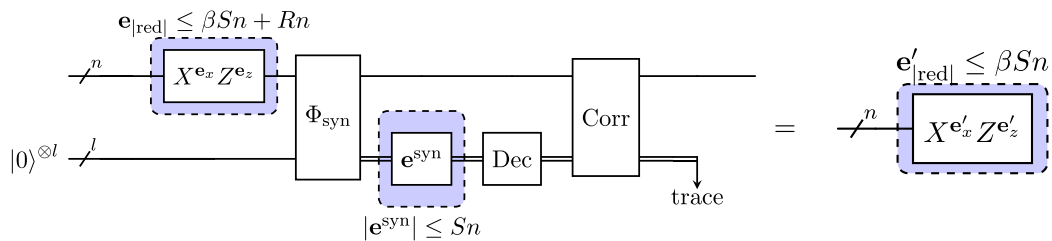


FIG. 4. The single-shot error correction procedure of Ref. [17]. The input is a code state of a QLDPC code, subjected to a Pauli error $X^{\mathbf{e}_x} Z^{\mathbf{e}_z}$. Then, the noiseless syndrome extraction circuit is applied, after which a bit-flip error $\mathbf{e}^{\text{syn}}$ affects the syndrome bits. Following classical decoding and error correction, the remaining error on the code state is $X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z}$. There exist constants $R > 0$, $S > 0$, and $\beta > 0$ with $\beta S n + R n < d_{\min}/2$, such that if the stabilizer reduced weight of the data error satisfies $|\mathbf{e}|_{\text{red}} \leq \beta S n + R n$ and the syndrome error satisfies $|\mathbf{e}^{\text{syn}}| \leq S n$, then the residual error satisfies $|\mathbf{e}'|_{\text{red}} \leq \beta S n$. Thus, the logical error is effectively suppressed using single-shot error correction.

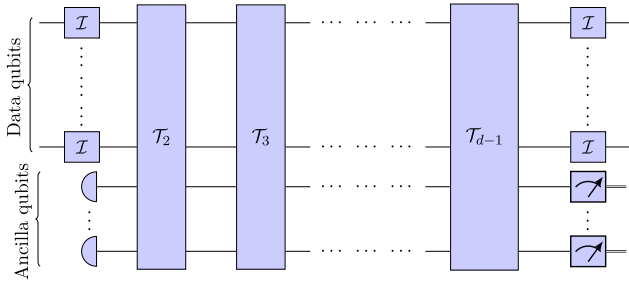


FIG. 5. The channel $\mathcal{T}_\Phi$ realized by the syndrome extraction circuit Φ . In the first layer, $i = 1$, the ancilla qubits are initialized and in the last layer, $i = d$, they are measured. In layers $i = 2, \dots, d-1$, the data and ancilla qubits are entangled, using two qubit gates.

$|\mathbf{e}^{\text{syn}}| \leq Sn$. Then,

$$U_{\text{ec}}(|\sigma\rangle \otimes (\Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}} X^{\mathbf{e}_x} Z^{\mathbf{e}_z} |\psi\rangle)) = \alpha_{\sigma + \mathbf{e}^{\text{syn}}, \sigma_{\mathbf{e}}} |\sigma\rangle \otimes X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z} |\psi\rangle, \quad (38)$$

where $\sigma \in \{0, 1\}^l$, the residual error $(\mathbf{e}'_x, \mathbf{e}'_z)$ satisfies $|\mathbf{e}'|_{\text{red}} \leq \beta Sn$, and the factor $\alpha_{\sigma + \mathbf{e}^{\text{syn}}, \sigma_{\mathbf{e}}}$ is defined as in Eq. (35).

Note that Eq. (38) can be stated using the encoding isometry $V: (\mathbb{C}^2)^{\otimes m} \rightarrow (\mathbb{C}^2)^{\otimes n}$:

$$U_{\text{ec}}(|\sigma\rangle \otimes (\Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}} X^{\mathbf{e}_x} Z^{\mathbf{e}_z} V)) = \alpha_{\sigma + \mathbf{e}^{\text{syn}}, \sigma_{\mathbf{e}}} |\sigma\rangle \otimes (X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z} V). \quad (39)$$

In the following corollary, we show that Theorem 4 can be generalized for arbitrary linear error operators.

Corollary 1 (Single-shot decoding for non-Pauli errors). There exist constants $R > 0$, $S > 0$, and $\beta > 0$ with $\beta Sn + Rn < d_{\min}/2$, such that the following holds:

Let $|\psi\rangle \in (\mathbb{C}^2)^{\otimes \bar{n}} \otimes \bar{L}$ be a quantum state on the joint space of the code space $\bar{L}$ and a $\bar{n}$ -qubit auxiliary system. Suppose that the data error is $E \in \mathbf{L}((\mathbb{C}^2)^{\otimes \bar{n}} \otimes (\mathbb{C}^2)^{\otimes n})$, with the reduced weight

$$|E|_{\text{red}} \leq \beta Sn + Rn < \frac{d_{\min}}{2},$$

and the syndrome error is $\mathbf{e}^{\text{syn}} \in \{0, 1\}^l$, $|\mathbf{e}^{\text{syn}}| \leq Sn$. Then,

$$(\mathbb{1}_{\bar{n}} \otimes U_{\text{ec}})(|\sigma\rangle \otimes (\mathbb{1}_{\bar{n}} \otimes \Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}}) E |\psi\rangle) = |\sigma\rangle \otimes (E_{\bar{n}, \sigma} \otimes X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z}) |\psi\rangle, \quad (40)$$

where $\sigma \in \{0, 1\}^l$, $E_{\bar{n}} \in \mathbf{L}((\mathbb{C}^2)^{\otimes \bar{n}})$ and for $\mathbf{e}' = (\mathbf{e}'_x, \mathbf{e}'_z)$, $|\mathbf{e}'|_{\text{red}} \leq \beta Sn$.

Proof. Using $\mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z)$, we have

$$\begin{aligned} (\mathbb{1}_{\bar{n}} \otimes \Pi_{\sigma}) E |\psi\rangle &= \sum_{\mathbf{e}} (E_{\mathbf{e}} \otimes \Pi_{\sigma} X^{\mathbf{e}_x} Z^{\mathbf{e}_z}) |\psi\rangle \\ &= \sum_{\mathbf{e}: \sigma_{\mathbf{e}} = \sigma} (E_{\mathbf{e}} \otimes \Pi_{\sigma} X^{\mathbf{e}_x} Z^{\mathbf{e}_z}) |\psi\rangle, \end{aligned} \quad (41)$$

where the first equality uses the expansion of E according to Eq. (32), and the second equality uses Eq. (35). Note that two errors with the same syndrome are separated by either a stabilizer generator or a logical operator. The reduced error weight of a logical operator is at least the minimum distance $d_{\min}$ of the code. Since $|E|_{\text{red}} < d_{\min}/2$, all the errors $\{\mathbf{e}: \sigma_{\mathbf{e}} = \sigma\}$ in Eq. (41) are stabilizer equivalent to each other. Therefore, Eq. (41) reduces to

$$(\mathbb{1}_{\bar{n}} \otimes \Pi_{\sigma}) E |\psi\rangle = (E_{\bar{n}, \sigma} \otimes \Pi_{\sigma} X^{\mathbf{e}_x} Z^{\mathbf{e}_z}) |\psi\rangle, \quad (42)$$

where $\mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z)$, $|\mathbf{e}|_{\text{red}} \leq \beta Sn + Rn$ is such that $\sigma_{\mathbf{e}} = \sigma$, and $E_{\bar{n}, \sigma} := \sum_{\mathbf{e}: \sigma_{\mathbf{e}} = \sigma} E_{\mathbf{e}}$. We now have

$$\begin{aligned} (\mathbb{1}_{\bar{n}} \otimes U_{\text{ec}})(|\sigma\rangle \otimes (\mathbb{1}_{\bar{n}} \otimes \Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}}) E |\psi\rangle) &= (\mathbb{1}_{\bar{n}} \otimes U_{\text{ec}})(|\sigma\rangle \otimes (E_{\bar{n}, \sigma + \mathbf{e}^{\text{syn}}} \otimes \Pi_{\sigma \oplus \mathbf{e}^{\text{syn}}} X^{\mathbf{e}_x} Z^{\mathbf{e}_z}) |\psi\rangle) \\ &= |\sigma\rangle \otimes (E_{\bar{n}, \sigma + \mathbf{e}^{\text{syn}}} \otimes X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z}) |\psi\rangle, \end{aligned}$$

where in the first equality, the syndrome of $\mathbf{e} = (\mathbf{e}_x, \mathbf{e}_z)$ is $\sigma_{\mathbf{e}} = \sigma \oplus \mathbf{e}^{\text{syn}}$ from Eq. (42) (note that U_{ec} acts on the syndrome state and the code block) and in the second equality, $|\mathbf{e}'|_{\text{red}} \leq \beta Sn$, using Theorem 4 [see also Eq. (38)]. ■

B. Single-shot error correction against general circuit noise

In this section, we consider the syndrome extraction circuit of a QLDPC code under general circuit noise with parameter δ . We show that single-shot error correction, i.e., the error correction based on a single round of noisy syndrome, suppresses the input error on any code state, given that δ is sufficiently small (see Theorem 5 below).

We know that for QLDPC codes, there exists a syndrome extraction circuit of constant (i.e., independent of n) depth [6]. Throughout this section, we denote by Φ the syndrome extraction circuit. Let $d = O(1)$ be the depth of Φ . The syndrome extraction circuit Φ contains the following steps (see also Fig. 5):

1. *Initialization.* The input to the circuit Φ corresponds to the n data (physical) qubits of the QLDPC code. In the first layer, $i = 1$, of the quantum circuit, we initialize l ancilla qubits (one for each generator measurement) in the $|0\rangle$ or $|+\rangle$ state.
2. *Entanglement.* In layers $i = 2, \dots, d-1$, we entangle the data qubits with ancilla qubits by applying CNOT gates.

3. *Measurement.* In the last layer, $i = d$, the ancilla qubits are measured in the Pauli-Z or X basis.

Note that for a QLDPC code, with at least one X - and Z -type generator of weight bigger than or equal to 2, it holds that $d \geq 4$. We denote the set of data qubits by D , and the set of ancilla qubits by A . The quantum channel realized by the circuit Φ is given by

$$\mathcal{T}_\Phi := \mathcal{T}_d \circ \cdots \circ \mathcal{T}_2 \circ \mathcal{T}_1, \quad (43)$$

where the quantum channel realized by the i^{th} layer is given by

$$\mathcal{T}_i := \otimes_{j=1}^{t_i} \mathcal{T}_{ij}, \quad i \in [d],$$

with $\mathcal{T}_{ij}$ being the channel corresponding to the location (i, j) , and t_i being the number of locations (including idle gates) at the i^{th} layer (see also Definition 4). Note that the following holds:

$$n \leq t_i < 2n, \quad \forall i \in [d]. \quad (44)$$

In Eq. (44), the upper bound follows using that the number of locations at any layer cannot exceed $n + l < 2n$ (recall that $l < n$ is the number of generator measurements). The lower bound follows from the observation that the maximum number of two-qubit locations at any layer is l , since each such location acts on a data and an ancilla qubit.

Applying general circuit noise with parameter δ on Φ , we obtain the set of quantum channels $\text{Trans}(\Phi, \delta)$ (see also Sec. IIC). Any $\tilde{\mathcal{T}}_\Phi \in \text{Trans}(\Phi, \delta)$ can be written as

$$\tilde{\mathcal{T}}_\Phi = \tilde{\mathcal{T}}_d \circ \cdots \circ \tilde{\mathcal{T}}_2 \circ \tilde{\mathcal{T}}_1, \quad (45)$$

where

$$\tilde{\mathcal{T}}_i = \otimes_{j=1}^{t_i} \tilde{\mathcal{T}}_{ij},$$

with $\tilde{\mathcal{T}}_{ij}$ such that $\|\tilde{\mathcal{T}}_{ij} - \mathcal{T}_{ij}\|_\diamond \leq \delta$.

In the following, we will use the notation $\mathcal{T}_{[l]} := \mathcal{T}_l \circ \cdots \circ \mathcal{T}_2 \circ \mathcal{T}_1$ and $\tilde{\mathcal{T}}_{[l]} := \tilde{\mathcal{T}}_l \circ \cdots \circ \tilde{\mathcal{T}}_2 \circ \tilde{\mathcal{T}}_1$, for $1 < i \leq d - 1$.

In Theorem 5 below, we consider parallel error correction across multiple code blocks of the QLDPC code, as needed in the proof of Theorem 3. We show that if the reduced weight of the input error is bounded in each code block, the single-shot error correction suppresses the input error in each code block, even with general circuit noise applied to the syndrome extraction circuit. Consequently, errors can be suppressed in all the code blocks in parallel. It is worth emphasizing that the input error can be correlated across different code blocks; the only requirement of Theorem 5 is that its weight with respect to each code block is bounded.

Theorem 5. There exist constants R, S , and $\beta > 0$, satisfying $\beta Sn + Rn < d_{\min}/2$, such that the following holds.

Let δ be the error rate, satisfying $5\delta 2^{d+1} \leq S$, with d being the depth of the syndrome extraction circuit Φ . Let $\kappa := ed$ with $e := \exp(1)$, and h be a positive integer, such that $h\kappa e \exp(-n\delta/3) \leq 1$. Consider parallel error correction of the h code blocks of the QLDPC code C . For the i^{th} code block, let (D_i, A_i) denote the data and ancilla qubits, $\bar{\mathcal{L}}^i$ denote the code space and let Φ^i denote the syndrome extraction circuit. Let $\mathcal{V}$ be a superoperator with the reduced weight $|\mathcal{V}|_{\text{red}} \leq \beta Sn + Rn$ with respect to all the code blocks $i \in [h]$, and let $|\psi\rangle \in \otimes_i \bar{\mathcal{L}}^i$ be any code state. Then, for any $\tilde{\mathcal{T}}_{\Phi^i} \in \text{Trans}(\Phi^i, \delta)$, $i \in [h]$,

$$\begin{aligned} & \left\| \left(\otimes_{i \in [h]} (\text{Tr}_{A_i} \circ \mathcal{U}_{\text{ec}}^i \circ \tilde{\mathcal{T}}_{\Phi^i}) \right) (\mathcal{V}(|\psi\rangle\langle\psi|)) \right. \\ & \quad \left. - \mathcal{V}(|\psi\rangle\langle\psi|) \right\|_1 < h\kappa e \exp\left(-\frac{n\delta}{3}\right), \end{aligned} \quad (46)$$

where $\mathcal{U}_{\text{ec}}^i$ corresponds to the error correction channel on the i^{th} block, Tr_{A_i} refers to tracing out ancilla qubits in A_i , and $\mathcal{V}'$ is a superoperator on h code blocks, satisfying $|\mathcal{V}'|_{\text{red}} \leq \beta Sn + 5n\delta 2^{d+1}$, with respect to all the code blocks $i \in [h]$.

Note that if $\delta < R/(5 \times 2^{d+1})$, then $|\mathcal{V}'|_{\text{red}} < |\mathcal{V}|_{\text{red}}$ in Theorem 5. Moreover, given that $h = \text{poly}(n)$, the right-hand side of Eq. (46) approaches zero as n goes to infinity. Therefore, the error weight can be effectively suppressed from $\beta Sn + Rn$ to $\beta Sn + 5n\delta 2^{d+1}$, by choosing sufficiently large n .

Before proving Theorem 5, we provide Lemma 2, which shows that for syndrome extraction circuit Φ , any $\tilde{\mathcal{T}}_\Phi \in \text{Trans}(\Phi, \delta)$ can be approximated by a superoperator $\bar{\mathcal{T}}_\Phi = \mathcal{T}_{\Phi_d} \circ \mathcal{N} \circ \mathcal{T}_{\Phi_{[d-1]}}$, where $\mathcal{N}$ is a superoperator of weight $O(n\delta)$. Therefore, the errors that happen in the circuit Φ can be pushed just before the single-qubit measurements are applied on the ancilla qubits. Intuitively, the weight of $\mathcal{N}$ is $O(n\delta)$ because the constant depth of the circuit Φ limits the propagation of errors.

Lemma 2. For any n such that $\exp(-n\delta/3) \leq 1/\kappa$, $\kappa = ed$, and any $\tilde{\mathcal{T}}_\Phi \in \text{Trans}(\Phi, \delta)$, there exists a superoperator $\bar{\mathcal{T}}_\Phi$ such that

$$\|\tilde{\mathcal{T}}_\Phi - \bar{\mathcal{T}}_\Phi\|_\diamond < \kappa \exp(-n\delta/3) \quad (47)$$

and

$$\bar{\mathcal{T}}_\Phi = \mathcal{T}_d \circ \mathcal{N} \circ \mathcal{T}_{[d-1]}, \quad (48)$$

where $\mathcal{N}$ is a superoperator of weight $5n\delta 2^{d+1}$.

Proof. We consider the first $d - 1$ layers of Φ (i.e., $\mathcal{T}_{[d-1]} = \mathcal{T}_{d-1} \circ \cdots \circ \mathcal{T}_2 \circ \mathcal{T}_1$) and the last layer (i.e., $\mathcal{T}_d$), separately (see also Fig. 5). We approximate any $\tilde{\mathcal{T}}_{[d-1]}$

by a superoperator $\bar{\mathcal{T}}_{[d-1]} = \mathcal{N}_{[d-1]} \circ \mathcal{T}_{[d-1]}$, and $\tilde{\mathcal{T}}_d$ by a superoperator $\bar{\mathcal{T}}_d = \mathcal{T}_d \circ \mathcal{N}_d$, where the weights of $\mathcal{N}_{[d-1]}$ and $\mathcal{N}_d$ are given by $5n\delta(2^{d+1} - 4)$, and $10n\delta$, respectively. We then take $\bar{\mathcal{T}}_\Phi := \bar{\mathcal{T}}_d \circ \bar{\mathcal{T}}_{[d-1]}$, which satisfies Eqs. (47) and (48).

For each layer i of the circuit, we do the following expansion:

$$\begin{aligned} \tilde{\mathcal{T}}_i &= \otimes_{j=1}^{t_i} (\mathcal{T}_{ij} + (\tilde{\mathcal{T}}_{ij} - \mathcal{T}_{ij})) \\ &= \sum_{B \subseteq [t_i]} (\otimes_{j \in [t_i] \setminus B} \mathcal{T}_{ij}) \otimes (\otimes_{j \in B} (\tilde{\mathcal{T}}_{ij} - \mathcal{T}_{ij})) \end{aligned}$$

For $s > 0$, we define a superoperator $\bar{\mathcal{T}}_i$ of weight s , as follows:

$$\bar{\mathcal{T}}_i := \sum_{B \subseteq [t_i], |B| \leq s} (\otimes_{j \in [t_i] \setminus B} \mathcal{T}_{ij}) \otimes (\otimes_{j \in B} (\tilde{\mathcal{T}}_{ij} - \mathcal{T}_{ij})) \quad (49)$$

Note that

$$\|\tilde{\mathcal{T}}_i - \bar{\mathcal{T}}_i\|_\diamond \leq \sum_{j > s} \binom{t_i}{j} \delta^j. \quad (50)$$

Then, by doing a similar calculation as in Lemma 1, we have that for $s = 5t_i\delta$ (using $t_i \geq n$),

$$\|\tilde{\mathcal{T}}_i - \bar{\mathcal{T}}_i\|_\diamond \leq \exp\left(-\frac{t_i\delta}{3}\right) \leq \exp\left(-\frac{n\delta}{3}\right). \quad (51)$$

For $\bar{\mathcal{T}}_{[d-1]} = \bar{\mathcal{T}}_{d-1} \circ \dots \circ \bar{\mathcal{T}}_1$, we have

$$\begin{aligned} \|\tilde{\mathcal{T}}_{[d-1]} - \bar{\mathcal{T}}_{[d-1]}\|_\diamond &\leq \|\tilde{\mathcal{T}}_{[d-1]} - \bar{\mathcal{T}}_{d-1} \circ \tilde{\mathcal{T}}_{[d-2]}\|_\diamond + \|\bar{\mathcal{T}}_{d-1} \circ \tilde{\mathcal{T}}_{[d-2]} - \bar{\mathcal{T}}_{[d-1]}\|_\diamond \\ &\leq \|\tilde{\mathcal{T}}_{d-1} - \bar{\mathcal{T}}_{d-1}\|_\diamond \|\tilde{\mathcal{T}}_{[d-2]}\|_\diamond + \|\bar{\mathcal{T}}_{d-1}\|_\diamond \|\tilde{\mathcal{T}}_{[d-2]} - \bar{\mathcal{T}}_{[d-2]}\|_\diamond \\ &\leq \exp\left(-\frac{n\delta}{3}\right) + \left(1 + \exp\left(-\frac{n\delta}{3}\right)\right) \|\tilde{\mathcal{T}}_{[d-2]} - \bar{\mathcal{T}}_{[d-2]}\|_\diamond, \end{aligned}$$

where the first inequality uses the triangle inequality, the second inequality uses $\|\mathcal{T}_2 \circ \mathcal{T}_1\|_\diamond \leq \|\mathcal{T}_2\|_\diamond \|\mathcal{T}_1\|_\diamond$, and the last inequality uses $\|\tilde{\mathcal{T}}_{[d-2]}\|_\diamond = 1$ and $\|\bar{\mathcal{T}}_{d-1}\|_\diamond \leq \|\bar{\mathcal{T}}_{d-1} - \tilde{\mathcal{T}}_{d-1}\|_\diamond + \|\tilde{\mathcal{T}}_{d-1}\|_\diamond \leq (\exp(-\frac{n\delta}{3}) + 1)$. Applying

the above recursively, we obtain

$$\begin{aligned} \|\tilde{\mathcal{T}}_{[d-1]} - \bar{\mathcal{T}}_{[d-1]}\|_\diamond &\leq \left(1 + \left(1 + \exp\left(-\frac{n\delta}{3}\right)\right) + \dots + \left(1 + \exp\left(-\frac{n\delta}{3}\right)\right)^{d-2}\right) \exp\left(-\frac{n\delta}{3}\right) \\ &< e(d-1) \exp\left(-\frac{n\delta}{3}\right), \end{aligned} \quad (52)$$

where in the second inequality, we have used $\exp(-n\delta/3) \leq 1/ed < 1/d$, which implies $\ln(1 + \exp(-n\delta/3)) < 1/d$, and in turn, $(1 + \exp(-n\delta/3))^d < e$.

Note from Eq. (49) that $\bar{\mathcal{T}}_i$ can be written as

$$\bar{\mathcal{T}}_i = \sum_{B \subseteq [t_i], |B| \leq s} (-1)^{|B|-|F|} \sum_{F \subseteq B} (\otimes_{j \in [t_i] \setminus F} \mathcal{T}_{ij}) \otimes (\otimes_{j \in F} \tilde{\mathcal{T}}_{ij}). \quad (53)$$

In other words, $\bar{\mathcal{T}}_i$ can be written as a linear combination of the superoperators

$$(\otimes_{j \in [t_i] \setminus F} \mathcal{T}_{ij}) \otimes (\otimes_{j \in F} \tilde{\mathcal{T}}_{ij}), \quad |F| \leq s = 5t_i\delta. \quad (54)$$

Since $\mathcal{T}_{ij}$, $\forall i \in [d-1]$ corresponds to either a unitary or a state preparation (all the measurements are performed at the last layer), we have that

$$\tilde{\mathcal{T}}_{ij} = \mathcal{N}_{ij} \circ \mathcal{T}_{ij}, \quad (55)$$

where $\mathcal{N}_{ij}$ is a quantum channel. Note that $\mathcal{N}_{ij} = \tilde{\mathcal{T}}_{ij} \circ \mathcal{T}_{ij}^\dagger$ if $\mathcal{T}_{ij}$ is a unitary and $\mathcal{N}_{ij} = \tilde{\mathcal{T}}_{ij} \circ \text{Tr}$ if $\mathcal{T}_{ij}$ is a state preparation channel, preparing a single-qubit state. Let F' be the set of qubits on which channels $\otimes_{j \in F} \mathcal{T}_{ij}$ act. Note that $|F'| \leq 2|F|$ as Φ contains only single- and two-qubit gates. Let $\mathcal{N}_{i,F'} := (\otimes_{j \in [t_i] \setminus F} \mathcal{T}_{ij}) \otimes (\otimes_{j \in F} \mathcal{N}_{ij})$, which has weight $|F'| \leq 2|F| \leq 10t_i\delta$. We now have

$$(\otimes_{j \in [t_i] \setminus F} \mathcal{T}_{ij}) \otimes (\otimes_{j \in F} \tilde{\mathcal{T}}_{ij}) = \mathcal{N}_{i,F'} \circ (\otimes_{j \in [t_i]} \mathcal{T}_{ij}) \quad (56)$$

$$= \mathcal{N}_{i,F'} \circ \mathcal{T}_i. \quad (57)$$

Therefore, we have $\bar{\mathcal{T}}_i = \mathcal{N}_i \circ \mathcal{T}_i$, where $\mathcal{N}_i$ is a superoperator of weight $10t_i\delta$ (as it is a linear combination of channels $\mathcal{N}_{i,F'}$).

Consider the set of qubits F' at the layer i . For $j > i$, let $\text{Shade}(F', j)$ be the set of qubits after the j th layer of the quantum circuit Φ that depend on the qubits in F' . In other words, $\text{Shade}(F', j)$ is the set of qubits in the future light cone of the set F' after applying layers $i+1, \dots, j$.

As Φ contains only single- and two-qubit gates, we have $\text{Shade}(F', i+1) \leq 2|F'|$. Therefore,

$$\begin{aligned} |\text{Shade}(F', j)| &\leq 2^{j-i}|F'| \leq 2^{j-i+1}|F| \\ &\leq 2^{j-i+1}5t_i\delta < 5n\delta 2^{j-i+2}, \end{aligned} \quad (58)$$

where the second inequality uses $|F'| \leq 2|F|$ and the last inequality uses $t_i \leq n + l < 2n$.

Using the unitarity of $\mathcal{T}_i, i \in [d-1]$, we can push $\mathcal{N}_{i,F'}$ to the left as follows:

$$\begin{aligned} &\mathcal{T}_{d-1} \circ \cdots \circ \mathcal{T}_{i+1} \circ (\mathcal{N}_{i,F'} \circ \mathcal{T}_i) \circ \cdots \circ \mathcal{T}_1 \\ &= \mathcal{T}_{d-1} \circ \cdots \circ \mathcal{N}'_{i,\text{Shade}(F',i+1)} \circ \mathcal{T}_{i+1} \circ \mathcal{T}_i \circ \cdots \circ \mathcal{T}_1 \\ &= \mathcal{N}'_{i,\text{Shade}(F',d-1)} \circ \mathcal{T}_{[d-1]}, \end{aligned} \quad (59)$$

where in the first equality $\mathcal{N}'_{i,\text{Shade}(F',i+1)} = \mathcal{T}_{i+1} \circ \mathcal{N}_{i,F'} \circ \mathcal{T}_{i+1}^\dagger$ acts trivially except on $\text{Shade}(F', i+1)$. Similarly, in the second equality $\mathcal{N}'_{i,\text{Shade}(F',d-1)}$ is a quantum channel that acts nontrivially only on $\text{Shade}(F', d-1)$, where $|\text{Shade}(F', d-1)| \leq 2^{d-i+1}5n\delta$ from Eq. (58). We now apply Eq. (59) for each layer $i = d-1, \dots, 1$ in that order and obtain

$$\bar{\mathcal{T}}_{[d-1]} = \mathcal{N}_{[d-1]} \circ \mathcal{T}_{[d-1]}, \quad (60)$$

where $\mathcal{N}_{[d-1]}$ is a linear combination of quantum channels that act nontrivially on at most $5n\delta(2^2 + 2^3 + \cdots + 2^d) = 5n\delta(2^{d+1} - 4)$ qubits. Therefore, the weight of $\mathcal{N}$ is given by $5n\delta(2^{d+1} - 4)$.

We now consider the d th layer of the circuit Φ . We have $\mathcal{T}_d = (\otimes_{j \in D} \mathcal{I}_j) \otimes (\otimes_{j \in A} \mathcal{M}_j)$, where $\mathcal{M}_j$ corresponds to a single qubit measurement in the Pauli-Z or -X basis (recall that D is the set of data qubits with $|D| = n$ and A is the set of ancilla qubits with $|A| = l$). Now consider $\tilde{\mathcal{T}}_d$ and the corresponding approximation $\bar{\mathcal{T}}_d$ from Eq. (49), obtained by using $t_d = n + l$,

$$\begin{aligned} \bar{\mathcal{T}}_d &= \sum_{\substack{B_1 \subseteq D, B_2 \subseteq A \\ |B_1| + |B_2| \leq 5(n+l)\delta}} (\otimes_{j \in D \setminus B_1} \mathcal{I}_j) \otimes (\otimes_{j \in A \setminus B_2} \mathcal{M}_j) \\ &\quad \otimes (\otimes_{j \in B_1} (\tilde{\mathcal{I}}_j - \mathcal{I}_j)) \otimes (\otimes_{j \in B_2} (\tilde{\mathcal{M}}_j - \mathcal{M}_j)), \end{aligned}$$

which satisfies

$$\|\tilde{\mathcal{T}}_d - \bar{\mathcal{T}}_d\|_\diamond \leq \exp\left(-\frac{n\delta}{3}\right). \quad (61)$$

The superoperator $\bar{\mathcal{T}}_d$ can be written as the linear combination of the following channels for a subset of qubits $F_1 \subseteq D, F_2 \subseteq A$, such that $|F_1| + |F_2| \leq 5(n+l)\delta < 10n\delta$:

$$(\otimes_{j \in D \setminus F_1} \mathcal{I}_j) \otimes (\otimes_{j \in A \setminus F_2} \mathcal{M}_j) \otimes (\otimes_{j \in F_1} \tilde{\mathcal{I}}_j) \otimes (\otimes_{j \in F_2} \tilde{\mathcal{M}}_j). \quad (62)$$

Consider single-qubit measurement $\mathcal{M}_j : \mathbf{L}(\mathbb{C}^2) \rightarrow \mathbb{B}$ in the computational basis defined by $\mathcal{M}_j(\cdot) = \sum_{a \in \{0,1\}}$

$\text{Tr}(|a\rangle\langle a| \cdot) |a\rangle\langle a|$. The noisy measurement $\tilde{\mathcal{M}}_j : \mathbf{L}(\mathbb{C}^2) \rightarrow \mathbb{B}$ corresponds to a quantum channel

$$\tilde{\mathcal{M}}_j(\cdot) = \sum_{a \in \{0,1\}} \text{Tr}(P_a \cdot) |a\rangle\langle a|, \quad (63)$$

where $P_a \geq 0, \sum_a P_a = \mathbb{1}$. It can be seen that

$$\tilde{\mathcal{M}}_j = \mathcal{M}_j \circ \mathcal{N}_j, \quad (64)$$

where the channel $\mathcal{N}_j$ is defined using qubit systems A and A' as below (see also Naimark's theorem [29, Theorem 2.42])

$$\begin{aligned} \mathcal{N}_j &= \text{Tr}_A(W_{A \rightarrow AA'}(\cdot)W_{A \rightarrow AA'}^\dagger), \quad W_{A \rightarrow AA'} \\ &:= \sum_{a \in \{0,1\}} \sqrt{P_a} \otimes |a\rangle_{A'}. \end{aligned} \quad (65)$$

Note that $W_{A \rightarrow AA'}$ is an isometry. In a similar way, for Pauli-X basis measurement, we have $\tilde{\mathcal{M}}_j = \mathcal{M}_j \circ \mathcal{N}_j$ for a quantum channel $\mathcal{N}_j$. Therefore,

$$\begin{aligned} &(\otimes_{j \in D \setminus F_1} \mathcal{I}_j) \otimes (\otimes_{j \in A \setminus F_2} \mathcal{M}_j) \otimes (\otimes_{j \in F_1} \tilde{\mathcal{I}}_j) \otimes (\otimes_{j \in F_2} \tilde{\mathcal{M}}_j) \\ &= (\otimes_{j \in D} \mathcal{I}_j) \otimes (\otimes_{j \in A} \mathcal{M}_j) \circ \mathcal{N}_{d,F_1,F_2} \\ &= \mathcal{T}_d \circ \mathcal{N}_{d,F_1,F_2}, \end{aligned} \quad (66)$$

where $\mathcal{N}_{d,F_1,F_2} := (\otimes_{j \in D \setminus F_1} \mathcal{I}_j) \otimes (\otimes_{j \in A \setminus F_2} \mathcal{I}_j) \otimes (\otimes_{j \in F_1} \tilde{\mathcal{I}}_j) \otimes (\otimes_{j \in F_2} \tilde{\mathcal{N}}_j)$. Therefore, we have

$$\bar{\mathcal{T}}_d = \mathcal{T}_d \circ \mathcal{N}_d, \quad (67)$$

where $\mathcal{N}_d$ is a superoperator of weight $10n\delta$. Finally, from Eqs. (52) and (61), we obtain

$$\begin{aligned} &\|\tilde{\mathcal{T}}_\Phi - \bar{\mathcal{T}}_\Phi\|_\diamond \\ &\leq \|\bar{\mathcal{T}}_d\|_\diamond \|\tilde{\mathcal{T}}_{[d-1]} - \bar{\mathcal{T}}_{[d-1]}\|_\diamond + \|\tilde{\mathcal{T}}_d - \bar{\mathcal{T}}_d\|_\diamond \\ &\leq \left(1 + \exp\left(-\frac{n\delta}{3}\right)\right) e(d-1) \\ &\quad \exp\left(-\frac{n\delta}{3}\right) + \exp\left(-\frac{n\delta}{3}\right) \\ &< ed \exp\left(-\frac{n\delta}{3}\right), \end{aligned} \quad (68)$$

where the second line uses Eqs. (52) and (61) and $\|\bar{\mathcal{T}}_d\|_\diamond \leq (1 + \exp(-n\delta/3))$, and the third line uses $1 + \exp(-n\delta/3) \leq 1 + 1/(ed) < 1 + (1 - 1/e)/(d-1) = (d-1/e)/(d-1)$ (recall that $d > 1$).

Therefore, $\bar{\mathcal{T}}_\Phi = \mathcal{T}_d \circ \mathcal{N} \circ \mathcal{T}_{[d-1]}$, where $\mathcal{N} := \mathcal{N}_d \circ \mathcal{N}_{[d-1]}$ is a quantum channel of weight $5n\delta 2^{d+1}$. Hence, Eqs. (47) and (48) hold. ■

We now prove Theorem 5.

1. Proof of Theorem 5

Similarly to Eq. (52), using the triangle inequality, and the equality $\|\mathcal{T}_1 \otimes \mathcal{T}_2\|_\diamond = \|\mathcal{T}_1\|_\diamond \|\mathcal{T}_2\|_\diamond$ for superoperators $\mathcal{T}_1, \mathcal{T}_2$ [5], we obtain from Eq. (47)

$$\begin{aligned} & \|(\otimes_{i \in [h]} \tilde{\mathcal{T}}_{\Phi^i}) - (\otimes_{i \in [h]} \bar{\mathcal{T}}_{\Phi^i})\|_\diamond \\ & \leq \left(1 + (1 + \kappa \exp(-\frac{n\delta}{3})) + \dots \right. \\ & \quad \left. + (1 + \kappa \exp(-\frac{n\delta}{3}))^{h-1}\right) \kappa \exp(-\frac{n\delta}{3}) \\ & \leq h(1 + \kappa \exp(-\frac{n\delta}{3}))^h \kappa \exp(-\frac{n\delta}{3}) \\ & \leq h\kappa \exp(h\kappa \exp(-\frac{n\delta}{3})) \exp(-\frac{n\delta}{3}) \\ & < h\kappa e \exp(-\frac{n\delta}{3}), \end{aligned} \quad (69)$$

where in the second to last inequality, we have used $(1+x)^h \leq e^{xh}$, and the last inequality uses the assumption $h\kappa \exp(-n\delta/3) \leq 1/e < 1$. From Eq. (69), we have

$$\begin{aligned} & \left\| \left(\otimes_{i \in [h]} (\text{Tr}_{A_i} \circ \mathcal{U}_{\text{ec}}^i \circ \tilde{\mathcal{T}}_{\Phi^i}) \right) (\mathcal{V}(|\psi\rangle\langle\psi|)) \right. \\ & \quad \left. - \left(\otimes_{i \in [h]} (\text{Tr}_{A_i} \circ \mathcal{U}_{\text{ec}}^i \circ \bar{\mathcal{T}}_{\Phi^i}) \right) (\mathcal{V}(|\psi\rangle\langle\psi|)) \right\|_1 \\ & < h\kappa e \exp(-\frac{n\delta}{3}). \end{aligned} \quad (70)$$

We now show that for any superoperator $\mathcal{V}$, such that $|\mathcal{V}|_{\text{red}} \leq \beta Sn + Rn$ with respect to all the code blocks $i \in [h]$, we have

$$\left(\otimes_{i \in [h]} (\text{Tr}_{A_i} \circ \mathcal{U}_{\text{ec}}^i \circ \bar{\mathcal{T}}_{\Phi^i}) \right) (\mathcal{V}(|\psi\rangle\langle\psi|)) = \mathcal{V}'(|\psi\rangle\langle\psi|), \quad (71)$$

where $|\mathcal{V}'|_{\text{red}} \leq \beta Sn + 5n\delta 2^{d+1}$ is satisfied for all the code blocks $i \in [h]$. Then, from Eqs. (70) and (71), we obtain Eq. (46).

First, note that the first $d-1$ layers of the syndrome extraction circuit Φ realize an isometry on the data qubits

$$\mathcal{T}_{[d-1]}(\cdot) = U_{[d-1]}(\cdot \otimes |0\rangle\langle 0|^{\otimes l}) U_{[d-1]}^\dagger, \quad (72)$$

where $U_{[d-1]}$ is a unitary such that for any $|\phi\rangle \in (\mathbb{C}^2)^{\otimes n}$, we have

$$(\mathbb{1}_D \otimes \langle \sigma |_A) U_{[d-1]}(|\phi\rangle \otimes |0\rangle_A^{\otimes l}) = \Pi_\sigma |\phi\rangle, \quad (73)$$

where $\Pi_\sigma \in \mathbf{L}((\mathbb{C}^2)^{\otimes n})$ is the projector on the space corresponding to the syndrome $\sigma \in \{0, 1\}^l$. The last layer of Φ realizes single qubit Pauli-Z measurements on ancilla qubits; therefore, the corresponding quantum channel is given by

$$\mathcal{T}_d = (\otimes_{j \in D} \mathcal{I}_j) \otimes \sum_{\sigma \in \{0, 1\}^l} |\sigma\rangle\langle\sigma|(\cdot)|\sigma\rangle\langle\sigma|. \quad (74)$$

We now focus on error correction for a fixed code block $j \in [h]$. Consider data qubits D_j and the ancilla qubits A_j corresponding to this block. Recall from Lemma 2 that

$$\bar{\mathcal{T}}_{\Phi^j} = \mathcal{T}_d^j \circ \mathcal{N}^j \circ \mathcal{T}_{[d-1]}^j, \quad (75)$$

where the superoperator $\mathcal{N}^j$ acts on the joint system $D_j \cup A_j$ and its weight is given by $5n\delta 2^{d+1}$. Therefore, $\mathcal{N}^j$ can be written as a linear combination of superoperators of the form $W(\cdot)W'$, where W and W' are operators acting on the joint system $D_j \cup A_j$, and their weight is $5n\delta 2^{d+1}$ (see also Definition 2). Here, for simplicity, we replace $\mathcal{N}^j$ by $W(\cdot)W'$. Similarly, we replace $\mathcal{V}$ by $E(\cdot)E'$, where E and E' are the operators that act on the joint code blocks $\cup_{i \in [h]} D_i$, satisfying $|E|_{\text{red}}, |E'|_{\text{red}} \leq \beta Sn + Rn$, for any code block $i \in [h]$.

For an $\mathbf{e}_x \in \{0, 1\}^{n+l}$ indicating a Pauli- X error on the data and ancilla qubits in the j th code block, we define $\mathbf{e}_x := (\mathbf{e}_x|_{D_j}, \mathbf{e}_x|_{A_j})$, where $\mathbf{e}_x|_{D_j} \in \{0, 1\}^n$ indicates the X error on the data qubits and $\mathbf{e}_x|_{A_j} \in \{0, 1\}^l$ that on the ancilla qubits. Similarly, we use $\mathbf{e}_z := (\mathbf{e}_z|_{D_j}, \mathbf{e}_z|_{A_j})$ for Pauli- Z errors. For clarity, we use the following notation:

$$|U_{[d-1]}^j E \psi\rangle := \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes U_{[d-1]}^j \right) (E |\psi\rangle \otimes |0\rangle_{A_j}^{\otimes l}), \quad (76)$$

where $U_{[d-1]}^j$ is a copy of $U_{[d-1]}$ in Eqs. (72) and (73). Note that the quantum state $|U_{[d-1]}^j E \psi\rangle$ corresponds to the joint data-ancilla state after the $(d-1)$ th layer of syndrome extraction applied only to the j th code block. For any measurement outcome $\sigma \in \{0, 1\}^l$ corresponding to ancilla qubit measurements in the d th layer of the syndrome extraction circuit, we have

$$\begin{aligned}
& \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes (\mathbb{1}_{D_j} \otimes \langle \sigma |_{A_j}) W \right) |U_{[d-1]}^j E \psi\rangle \\
&= \sum_{\mathbf{e}_x, \mathbf{e}_z \in \{0,1\}^{n+l}} W_{\mathbf{e}_x, \mathbf{e}_z} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes (\mathbb{1}_{D_j} \otimes \langle \sigma |_{A_j}) Z^{\mathbf{e}_z} X^{\mathbf{e}_x} \right) |U_{[d-1]}^j E \psi\rangle \\
&= \sum_{\mathbf{e}_x, \mathbf{e}_z} W_{\mathbf{e}_x, \mathbf{e}_z} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes (Z^{\mathbf{e}_z|_{D_j}} X^{\mathbf{e}_x|_{D_j}} \otimes \langle \sigma |_{A_j} Z^{\mathbf{e}_z|_{A_j}} X^{\mathbf{e}_x|_{A_j}}) \right) |U_{[d-1]}^j E \psi\rangle \\
&= \sum_{\mathbf{e}_x, \mathbf{e}_z} (-1)^{\mathbf{e}_z|_{A_j} \cdot \sigma} W_{\mathbf{e}_x, \mathbf{e}_z} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes (Z^{\mathbf{e}_z|_{D_j}} X^{\mathbf{e}_x|_{D_j}} \otimes \langle \sigma |_{A_j} X^{\mathbf{e}_x|_{A_j}}) \right) |U_{[d-1]}^j E \psi\rangle \\
&= \sum_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} W_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes (X^{\mathbf{e}_x|_{D_j}} Z^{\mathbf{e}_z|_{D_j}} \otimes \langle \sigma \oplus \mathbf{e}_x |_{A_j}) \right) |U_{[d-1]}^j E \psi\rangle \\
&= \sum_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} W_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes X^{\mathbf{e}_x|_{D_j}} Z^{\mathbf{e}_z|_{D_j}} \Pi_{(\sigma \oplus \mathbf{e}_x|_{A_j})} \right) E |\psi\rangle, \tag{77}
\end{aligned}$$

where in the first equality, we expand the operator W in the Pauli basis, i.e., $W = \sum_{\mathbf{e}_x, \mathbf{e}_z \in \{0,1\}^{n+l}} W_{\mathbf{e}_x, \mathbf{e}_z} Z^{\mathbf{e}_z} X^{\mathbf{e}_x}$, and in the fourth equality, we have used $W_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} := \sum_{\mathbf{e}_z|_{A_j}} (-1)^{\mathbf{e}_z|_{A_j} \cdot \sigma} W_{\mathbf{e}_x, \mathbf{e}_z}$. In the last equality, we use Eq. (73). As the weight of W is $5n\delta 2^{d+1}$, it follows that

the Hamming weights of $\mathbf{e}_x|_{D_j}, \mathbf{e}_z|_{D_j}, \mathbf{e}_x|_{A_j}, \mathbf{e}_z|_{A_j}$ are upper bounded by $5n\delta 2^{d+1}$.

Applying the error correction U_{ec} on the data qubits corresponding to the j th code block, when the observed syndrome is $\sigma \in \{0,1\}^l$, we obtain

$$\begin{aligned}
& \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes U_{\text{ec}}(\mathbb{1}_{D_j} \otimes |\sigma\rangle\langle\sigma|_{A_j}) W \right) |U_{[d-1]}^j E \psi\rangle \\
&= \sum_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} W_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes U_{\text{ec}}(|\sigma\rangle_{A_j} \otimes Z^{\mathbf{e}_z|_{D_j}} X^{\mathbf{e}_x|_{D_j}} \Pi_{(\sigma \oplus \mathbf{e}_x|_{A_j})}) \right) E |\psi\rangle \\
&= \sum_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} W'_{\mathbf{e}_x, \mathbf{e}_z|_{D_j}} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes Z^{\mathbf{e}_z|_{D_j}} X^{\mathbf{e}_x|_{D_j}} U_{\text{ec}}(|\sigma\rangle_{A_j} \otimes \Pi_{(\sigma \oplus \mathbf{e}_x|_{A_j})}) \right) E |\psi\rangle \\
&= |\sigma\rangle_{A_j} \otimes \left(\sum_{\mathbf{e}_z|_{D_j}, \mathbf{e}_x} W'_{\mathbf{e}_x, \mathbf{e}_z|_{D_j}} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes Z^{\mathbf{e}_z|_{D_j}} X^{\mathbf{e}_x|_{D_j}} \right) (E_{\neq j} \otimes X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z}) |\psi\rangle \right), \tag{78}
\end{aligned}$$

where the first equality uses Eq. (77), in the second equality, we use that U_{ec} is a controlled Pauli, and $W'_{\mathbf{e}_x, \mathbf{e}_z|_{D_j}} = \pm W_{\mathbf{e}_x, \mathbf{e}_z|_{D_j}}$. In the last equality, from Corollary 1, $E_{\neq j} = 1/2^n \sum_{\bar{\mathbf{e}}: \sigma_{\bar{\mathbf{e}}} = \sigma \oplus \mathbf{e}_x|_{A_j}} \text{Tr}_{D_j}(E((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes Z^{\bar{\mathbf{e}}_z} X^{\bar{\mathbf{e}}_x}))$, and for $\mathbf{e} = (\mathbf{e}'_x, \mathbf{e}'_z) | \mathbf{e}'|_{\text{red}} \leq \beta S n$. Consider the total error operator in Eq. (78),

The stabilizer reduced weight of $\bar{E}$ with respect to the j th code block is upper bounded by $\beta S n + 5n\delta 2^{d+1}$. The reduced weight of $\bar{E}$ for code blocks $i \neq j$ remains upper bounded by $\beta S n + R n$. Similarly, after doing error correction with respect to all the code blocks, the residual error operator $\bar{E}$, satisfies $|\bar{E}|_{\text{red}} \leq \beta S n + 5n\delta 2^{d+1}$ with respect to all the code blocks $i \in [h]$. Similarly, Eq. (78) also holds for the error operators W' and E' ; therefore, it follows that

$$\begin{aligned}
\bar{E} &:= \sum_{\mathbf{e}_x, \mathbf{e}_z|_{D_j}} W'_{\mathbf{e}_x, \mathbf{e}_z|_{D_j}} \left((\otimes_{i \neq j} \mathbb{1}_{D_i}) \otimes X^{\mathbf{e}_x|_{D_j}} Z^{\mathbf{e}_z|_{D_j}} \right) (E_{\neq j} \otimes X^{\mathbf{e}'_x} Z^{\mathbf{e}'_z}) \\
& \tag{79} \\
& (\otimes_i (\text{Tr}_{A_i} \circ \mathcal{U}_{\text{ec}}^i \circ \bar{T}_{\Phi^i})) \circ \mathcal{V}(|\psi\rangle\langle\psi|) = \mathcal{V}'(|\psi\rangle\langle\psi|) \\
& := \sum_{\bar{E}, \bar{E}'} \bar{E} |\psi\rangle\langle\psi| \bar{E}', \tag{80}
\end{aligned}$$

where $|\mathcal{V}'|_{\text{red}} \leq \beta S n + 5n\delta 2^{d+1}$, with respect to all the code blocks. Finally, from Eq. (70) and (80), we obtain Eq. (46).

ACKNOWLEDGMENTS

We would like to thank Paula Belzig, Alexander Müller-Hermes, Robert König, Thomas Theurer, and Freek Witteveen for useful discussions. M.C. and A.G. acknowledge financial support from the European Research Council (ERC) (Grant Agreement No. 818761), the Villum Foundation via the QMATH Centre of Excellence (Grant No. 10059), and the Novo Nordisk Foundation (Grant No. NNF20OC0059939 “Quantum for Life”). Part of this work was completed while M.C. was Turing Chair for Quantum Software, associated to the QuSoft research center in Amsterdam, acknowledging financial support by the Dutch National Growth Fund (NGF), as part of the Quantum Delta NL visitor programme. O.F. acknowledges financial support from the ERC (Grant Agreement No. 851716) and from a government grant managed by the Agence Nationale de la Recherche under Plan France 2030 with the reference ANR-22-PETQ-0006. We also thank the National Center for Competence in Research SwissMAP of the Swiss National Science Foundation and the Section of Mathematics at the University of Geneva for their hospitality.

DATA AVAILABILITY

No data were created or analyzed in this study.

- [1] D. Gottesman in *Quantum Information Science and Its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics* (American Mathematical Society (AMS), 2010), Vol. 68, p. 13.
- [2] J. Preskill, in *Introduction to Quantum Computation and Information* (World Scientific, 1998), p. 213.
- [3] D. Aharonov and M. Ben-Or in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery (ACM), 1997), p. 176.
- [4] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, *Quantum Inf. Comput.* **6**, 97 (2006).
- [5] A. Yu. Kitaev, Quantum computations: Algorithms and error correction, *Russ. Math. Surv.* **52**, 1191 (1997).
- [6] D. Gottesman, Fault-tolerant quantum computation with constant overhead, *Quantum Inf. Comput.* **14**, 1338 (2014).
- [7] O. Fawzi, A. Grospellier, and A. Leverrier, Constant overhead quantum fault tolerance with quantum expander codes, *Commun. ACM* **64**, 106 (2020).
- [8] H. Yamasaki and M. Koashi, Time-efficient constant-space-overhead fault-tolerant quantum computation, *Nat. Phys.* **20**, 247 (2024).
- [9] Recently, we have become aware of two related works [30,31] showing that for local stochastic noise, the time overhead can be further reduced to polylogarithmic, while still maintaining constant qubit overhead.
- [10] J. P. Bonilla Ataides, H. Zhou, Q. Xu, G. Baranes, B. Li, M. D. Lukin, and L. Jiang, Constant-overhead fault-tolerant Bell-pair distillation using high-rate codes, *Phys. Rev. Lett.* **135**, 130804 (2025).
- [11] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, High-threshold and low-overhead fault-tolerant quantum memory, *Nature* **627**, 778 (2024).
- [12] A. Cowtan, Z. He, D. J. Williamson, and T. J. Yoder, Parallel logical measurements via quantum code surgery, *ArXiv:2503.05003*.
- [13] Q. Xu, J. Pablo Bonilla Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasić, M. D. Lukin, L. Jiang, and H. Zhou, Constant-overhead fault-tolerant quantum computation with reconfigurable atom arrays, *Nat. Phys.* **20**, 1084 (2024).
- [14] S. Yoshida, S. Tamiya, and H. Yamasaki, Concatenate codes, save qubits, *npj Quantum Inf.* **11**, 88 (2025).
- [15] A. Leverrier and G. Zémor, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), p. 872.
- [16] P. Panteleev and G. Kalachev, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery (ACM), 2022), p. 375.
- [17] S. Gu, E. Tang, L. Caha, S. H. Choe, Z. He, and A. Kubica, Single-shot decoding of good quantum LDPC codes, *Commun. Math. Phys.* **405**, 85 (2024).
- [18] M. Christandl, O. Fawzi, and A. Goswami, Fault-tolerant quantum input/output, *ArXiv:2408.05260*.
- [19] D. Kretschmann, D. Schlingemann, and R. F. Werner, The information-disturbance tradeoff and the continuity of Stinespring’s representation, *IEEE Trans. Inf. Theory* **54**, 1708 (2008).
- [20] J. Watrous, Semidefinite programs for completely bounded norms, *Theory Comput.* **5**, 217 (2009).
- [21] A. Grospellier, Ph.D. thesis, École doctorale Informatique, Télécommunications et Électronique, INRIA Paris, équipe SECRET, Sorbonne Université, Paris, 2019.
- [22] We have chosen the block size $m = O(\sqrt{x})$ for the simplicity of notation as in Ref. [21]. In principle, one can choose $m = O(x/\text{polylog}(x))$ [6].
- [23] D. Gottesman, Ph.D. thesis, California Institute of Technology, Pasadena, California, 1997.
- [24] It is possible to reduce the depth of Ψ'_{U_g} to $O(\log n)$ at the cost of extra ancillary qubits [32].
- [25] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A—At., Mol. Opt. Phys.* **86**, 032324 (2012).
- [26] P. W. Shor, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, 1996), p. 56.
- [27] H. Bombin, Single-shot fault-tolerant quantum error correction, *Phys. Rev. X* **5**, 031043 (2015).
- [28] E. T. Campbell, A theory of single-shot error correction for adversarial noise, *Quantum Sci. Technol.* **4**, 025006 (2019).
- [29] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).

- [30] Q. T. Nguyen and C. A. Pattison, Quantum fault tolerance with constant-space and logarithmic-time overheads, [ArXiv:2411.03632](#).
- [31] S. Tamiya, M. Koashi, and H. Yamasaki, Polylog-time- and constant-space-overhead fault-tolerant quantum computation with quantum low-density parity-check codes, [ArXiv:2411.03683](#).
- [32] C. Moore and M. Nilsson, Parallel quantum computation and quantum codes, [SIAM J. Comput.](#) **31**, 799 (2001).