

Efficient Self-Consistent Learning of Gate Set Pauli Noise

Senrui Chen^{1,2,*}, Zhihan Zhang³, Liang Jiang¹, and Steven T. Flammia^{4,5}

¹*Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA*

²*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California, USA*

³*Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China*

⁴*Department of Computer Science, Virginia Tech, Alexandria, Virginia, USA*

⁵*Phasecraft Inc., Washington, DC, USA*

 (Received 30 April 2025; revised 23 August 2025; accepted 5 November 2025; published 9 January 2026)

Understanding quantum noise is an essential step towards building practical quantum information processing systems. Pauli noise is a useful model that has been widely applied in quantum benchmarking, error mitigation, and error correction. Despite intensive study into Pauli noise learning, most existing works focus on learning Pauli channels associated with some specific gates rather than treating the gate set as a whole. A learning algorithm that is self-consistent, complete, and efficient at the same time has yet to be established. In this work, we study the task of *gate set Pauli noise learning*, where a set of quantum gates, state preparation, and measurements all suffer from unknown Pauli noise channels with a customized noise ansatz. Using tools from algebraic graph theory, we analytically characterize the self-consistently learnable degrees of freedom for Pauli noise models with arbitrary linear ansatz, and design experiments to efficiently learn all the learnable information. Specifically, we show that all learnable information about the gate noise can be learned to relative precision, under mild assumptions on the noise ansatz. We then demonstrate the flexibility of our theory by applying it to concrete physically motivated ansatzes (such as spatially local or quasi-local noise) and experimentally relevant gate sets (such as parallel CZ gates). These results not only enhance the theoretical understanding of quantum noise learning, but also provide a feasible recipe for characterizing existing and near-future quantum computing devices.

DOI: [10.1103/1pnv-t9px](https://doi.org/10.1103/1pnv-t9px)

I. INTRODUCTION

Quantum technologies are anticipated to offer significant advantages in various information processing tasks, including computation [1,2], communication [3,4], metrology [5,6], and more. Despite rapid experimental progress, noise remains as a major challenge in building a practical quantum information processing platform. In order to suppress, mitigate, and correct noise affecting a quantum device, it is essential to first gain knowledge about the noise. However, learning quantum noise is not easy. Noise is complicated, containing a large number of parameters that grows exponentially with the system size. Meanwhile, the available quantum control for noise learning is limited and can itself suffer from imperfection. Developing

efficient, reliable, and informative quantum noise characterization protocols is thus an important problem and a field of active research [7–9].

Among different noise characterization approaches, Pauli noise learning is a simple yet powerful formalism that has gained increasing attention recently [10–22]. Here, the noise processes affecting different components (i.e., state preparation, measurements, gates) of a quantum device are modeled by Pauli channels, defined as a stochastic mixture of multi-qubit Pauli operations [see Eq. (9)]. Pauli channels can naturally describe a wide spectrum of incoherent noises, including depolarizing and dephasing noises. It is also a widely used model in the study of quantum error correction and fault-tolerant quantum computing [23]. Besides, there exist techniques known as randomized compiling [24–26] that, given mild assumptions about the noise and gate set, can tailor an arbitrary noise channel into a Pauli channel, justifying the validity of this noise model.

Despite intensive recent study, some core aspects of Pauli noise learning have not been well-understood. In particular, a learning algorithm that is simultaneously *self-consistent*, *comprehensive*, and *efficient* has yet to be established. Here, self-consistency means using only

*Contact author: csenrui@gmail.com

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.

operations that are noisy to learn about themselves, comprehensiveness means to learn all the information that is self-consistently learnable, and efficiency means the number of measurements and classical processing timescales favorable with the number of qubits. In this work, we build up a framework called *gate set Pauli noise learning*, give analytic characterization of the self-consistently learnable degrees of freedom, and develop an algorithm with the above three desirable features.

Problem Setup. Consider an n -qubit quantum system. Suppose, by design, the following operations can be implemented on the platform: (1) Initialize the system to the ground state; (2) Apply a layer of n single-qubit unitary gates; (3) Apply some multi-qubit Clifford gates from a finite set denoted by $\mathcal{G}$; (4) Measure all qubits on the computational basis. We refer to the collection of all the above components as a *gate set*, following the language of gate set tomography (GST) [27]. Note that gate sets include noisy state preparation and measurement (SPAM). In practice, all the operations of a gate set can be noisy. Our goal is to characterize the noisy gate set *self-consistently*. That is, using only operations from the gate set to learn about themselves.

Instead of dealing with the most general form of noise, we focus on what is known as the *Pauli noise model*: We ignore the noise of any single-qubit gates, and assume the noise channels acting on the initial state, measurements, and any multi-qubit Clifford gate to be a (component-dependent) Pauli channel, which takes the following form:

$$\Lambda(\rho) = \sum_{a \in \mathbf{P}^n} p_a P_a \rho P_a = \sum_{b \in \mathbf{P}^n} \lambda_b P_b \text{Tr}(P_b \rho) / 2^n. \quad (1)$$

Here, $\mathbf{P}^n := \{I, X, Y, Z\}^{\otimes n}$ is the n -qubit Pauli group (modulo phase). $\{p_a\}$ are the *Pauli error rates* while $\{\lambda_b\}$ are the *Pauli fidelities* or *Pauli eigenvalues*. We assume $\lambda_b > 0$ throughout this work. Such a model can be ensured via randomized compiling under mild assumptions [25,26]. The action of noise channel on the initial state ρ_0 , positive operator-valued measure (POVM) element E_j , and gate $\mathcal{G} \in \mathcal{G}$ is given by

$$\tilde{\rho}_0 = \Lambda^S(\rho_0), \quad \tilde{E}_j = \Lambda^M(E_j), \quad \tilde{\mathcal{G}} = \Lambda^G \circ \mathcal{G}, \quad (2)$$

respectively. A realization of the gate set Pauli noise model is specified by the involved Pauli channels $\{\Lambda^S, \Lambda^M, \Lambda^G : \mathcal{G} \in \mathcal{G}\}$. Since a generic Pauli channel contains exponentially many parameters, we often need to assume an efficient parametrization of the Pauli channels, which we call a *noise ansatz*. Examples of noise ansatzes include spatially local or quasi-local Pauli channels, as shown in Figs. 1(c) and 1(d), respectively. We refer to a Pauli noise model with a noise ansatz as a *reduced model*, and the one without a noise ansatz as a *complete model*. Throughout this work, we assume the true noise model is *realizable*. That is, the true noise model can be represented exactly as a point in the model space.

To characterize the noisy gate set means to learn all the Pauli noise channels with ansatzes. However, it is generally impossible to fully determine those Pauli noise channels self-consistently [18,28]. Indeed, consider the following *gauge transformation* [27] on the noisy

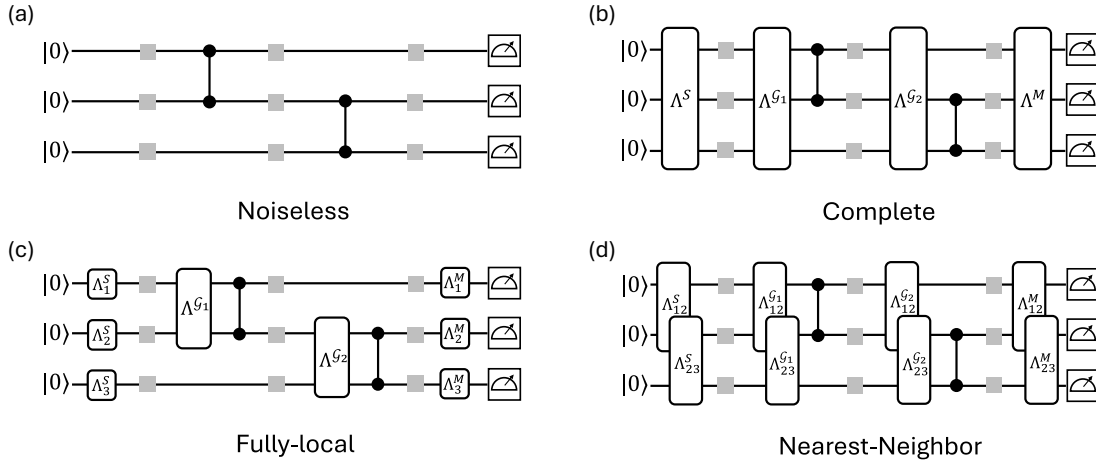


FIG. 1. Examples of different noise ansatz on a three-qubit system. The available set of multi-qubit gates are two CZ gates. The small square box represents single-qubit gates whose noise is ignored (or absorbed to multi-qubit gates and SPAM). (a) A noiseless circuit. (b) Circuit with complete Pauli noise, where the noise for SPAM and both gates are all generic three-qubit Pauli channels. (c) Circuit with fully local Pauli noise, where the SPAM noise channels are qubit-wise independent, while the gate noise channels have the same support as the gate. (d) Circuit with nearest-neighbor Pauli noise (a special case of the quasi-local model), where each Pauli noise channel is nearest-neighbor two-local. See Sec. VI for a rigorous definition.

operations,

$$\begin{aligned}\tilde{\rho}_0 &\mapsto \mathcal{D}(\tilde{\rho}_0), \quad \tilde{E}_j \mapsto \mathcal{D}^{-1}(\tilde{E}_j), \\ \tilde{\mathcal{G}} &\mapsto \mathcal{D}\tilde{\mathcal{G}}\mathcal{D}^{-1}, \quad \bigotimes_{i=1}^n \mathcal{U}_i \mapsto \mathcal{D} \left(\bigotimes_{i=1}^n \mathcal{U}_i \right) \mathcal{D}^{-1},\end{aligned}\quad (3)$$

where $\mathcal{D}$ is some invertible linear map. It is not hard to see that any experimental outcome statistics remain unchanged under such a transformation. Thus, two realizations of a noisy gate set related by a gauge transformation are *indistinguishable*. Meanwhile, one can carefully choose $\mathcal{D}$ (e.g., as a depolarizing channel on a subset of qubits) such that the transformed noisy gate set will remain a valid Pauli noise model within ansatz. That is, the single-qubit gates $\bigotimes_i \mathcal{U}_i$ remain noiseless, while $\Lambda^S, \Lambda^M, \Lambda^G$ remain Pauli channels within the ansatz (possibly with different parameters).

Now, consider a function of the gate set noise parameters. If the function value changes under certain valid gauge transformations, then the function is not self-consistently learnable. This raises the first question of gate set Pauli noise learning as follows:

Problem 1. How to classify the learnable and unlearnable functions of a gate set Pauli noise model?

This question is partially addressed in Ref. [18], but two aspects remain open. First, Ref. [18] only classifies learnable and unlearnable functions of gate-dependent noise parameters but not for SPAM noise parameters, making it incomplete in the sense of gate set characterization. Second, it only studies the unconstrained case of Pauli noise channels, while the classification under efficient Pauli noise ansatzes remains open.

After understanding what are the learnable degrees of freedom, the next natural problem is to find an efficient learning protocol, which can be summarized as follows:

Problem 2. How to efficiently learn all the learnable information of a gate set Pauli noise model?

In this work, we will establish a generic theoretic framework to address the above two problems. To demonstrate the power and flexibility of our framework, we will apply it to Pauli noise models with concrete gate set and noise ansatzes that are considered in the literature (e.g., Refs. [14,17]) and resolve open problems therein. A detailed summary of our contributions is provided in the next section.

A. Summary of results

1. Learnability of gate set Pauli noise

Our approach to characterize the learnability (i.e., classification of learnable and unlearnable information) is to

first encode all the noise parameters into a linear space. Let us start with the complete Pauli noise model (i.e., the one without a noise ansatz). Similar results have been obtained in Ref. [18] for this case, but here we will use a more formal way that can be extended to the reduced model later.

Define the logarithmic Pauli eigenvalues as $x_b := -\log \lambda_b$. We collect all logarithmic Pauli eigenvalues (excluding the trivial ones corresponding to identity operator) from all Pauli noise channels into an ordered list denoted by $\mathbf{x}$, which can be viewed as a vector of a real linear space according to a standard basis. We denote the space $\mathbf{x}$ lives on as X , called the *parameter space*. An *experiment* is defined as a mapping $X \mapsto \mathbb{R}^{2^n}$ specified by a sequence of gates $\mathcal{C}$. It maps $\mathbf{x}$ to a measurement outcome probability distribution given by the Born rule, $\Pr(j) = \text{Tr}[\tilde{E}_j \tilde{\mathcal{C}}(\tilde{\rho}_0)]$, where the noisy realization of state, gates, and measurement is determined by $\mathbf{x}$. Now, we define a linear function $\mathbf{f} \in \mathcal{L}(X \mapsto \mathbb{R}) =: X'$ to be *learnable* if its value can be completely specified by a set of experiments [29]; We also define a vector $\mathfrak{d} \in X$ to be a *gauge vector* if the two lists of noise parameters $\mathbf{x}$ and $\mathbf{x} + \mathfrak{d}$ cannot be distinguished by any experiment, for all $\mathbf{x} \in X$. The linear subspaces of all learnable linear function and all gauge vectors are denoted by L and T , respectively. The problem then becomes characterizing X , L , and T .

For this purpose, we introduce a directed graph called the Pauli pattern transfer graph (PTG). This graph is first defined in Ref. [18] with some twist needed to fit into the current framework. The graph consists of 2^n nodes. One of which is a root node (also called a SPAM node), the others correspond to the $2^n - 1$ non-zero n -bit strings, each denotes a Pauli pattern (the pattern of a n -qubit Pauli operator can be understood as its support); For each logarithmic Pauli eigenvalues, we assign a unique edge to the PTG: For gate noise parameters, each edge describe how a Clifford gate transform certain Pauli operators (where we only track the pattern); For each state preparation (or, measurement) noise parameter, we assign an edge from the root node to certain pattern node (or, from a pattern node to root). An example of PTG is given in Fig. 2.

PTG contains all information about a complete Pauli noise model. Intuitively, a walk starting from and ending at the root node gives a combination of logarithmic Pauli eigenvalues that can be learned from an experiment. This intuition can be made rigorous using tools from algebraic graph theory [30,31]. Informally, let the edge space E to be the real linear space spanned by all edges of PTG. For each cycle (or, cut) on PTG, we can assign it with a vector of E in a proper way (see Sec. III B). The subspace spanned by all cycle (or, cut) vectors is defined as the *cycle space* Z (or, the *cut space* U). It is known that Z and U are orthogonal complements to each other within E . Our main theorem is that, up to the natural identification between the parameter space X and the edge space E , the learnable space L is

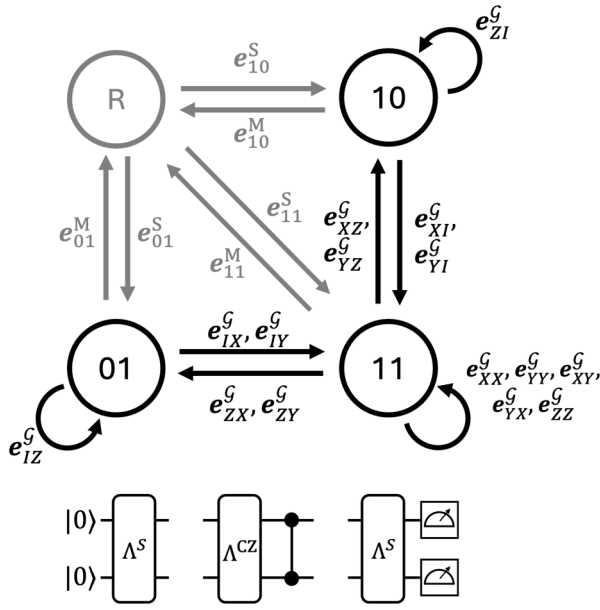


FIG. 2. The pattern transfer graph (PTG) for a two-qubit gate set $\mathcal{G} = \{CZ\}$. The gate set is depicted at the bottom of the graph. Here we define $\mathcal{G} = CZ$ for notational simplicity. The black edges correspond to gate noise parameters, while the gray edges correspond to SPAM noise parameters. Multiple edges are shown as one edge with multiple labels.

identical to the cycle space Z , while the gauge space T is identical to the cut space U . This implies that L and T are also orthogonal complements to each other within X [32].

Next, we turn to study the learnability of a reduced model, i.e., a noise model with ansatzes. A reduced model is specified by a tuple $(X_R, \mathcal{Q})$. Here X_R is the *reduced parameter space*, a linear space whose components are viewed as lists of the reduced noise parameters (denoted by $\mathbf{r}$). $\mathcal{Q} : X_R \mapsto X$ is the *embedding map* describing how the reduced parameters specifies the Pauli channels. Specifically, the logarithmic Pauli eigenvalues are given by $\mathbf{x} = \mathcal{Q}(\mathbf{r})$. We will assume $\mathcal{Q}$ is linear and injective [33]. The notion of experiments on the reduced model can be naturally induced from the definition for the complete model. We can also similarly define the reduced learnable space L_R (learnable functions of X_R) and the reduced gauge space T_R (indistinguishable transformations within X_R). The problem is then to characterize X_R , L_R , T_R . The challenge here is that, it is not easy to construct a graph describing the reduced noise parameters as in the complete model. Our approach is to connect the reduced spaces back to the complete spaces, then use the characterization that we have already obtained. We proved that $L_R = \mathcal{Q}^\top(L)$, $T_R = \mathcal{Q}^{-1}(T)$. In words, the reduced learnable space is the complete learnable space projected by $\mathcal{Q}^\top$, while the reduced gauge space is the preimage of the complete gauge space via $\mathcal{Q}$.

The relation between all the linear spaces introduced above can be summarized as follows.

Theorem 1 (Learnability, informal). For any reduced model $(X_R, \mathcal{Q})$ such that $\mathcal{Q}$ is linear and injective,

$$\begin{array}{lcl}
 \text{Reduced model:} & X_R & \\
 & \text{(reduced parameter)} & \\
 & \downarrow \mathcal{Q} & \\
 \text{Complete model:} & X & \\
 & \text{(parameter)} & \\
 & \parallel & \\
 \text{Pattern transfer graph:} & E & \\
 & \text{(edge)} &
 \end{array}
 \quad
 \begin{array}{lcl}
 = & L_R & \oplus^\perp T_R \\
 & \text{(reduced learnable)} & \text{(reduced gauge)} \\
 & \uparrow \mathcal{Q}^\top & \uparrow \mathcal{Q}^{-1} \\
 = & L & \oplus^\perp T \\
 & \text{(learnable)} & \text{(gauge)} \\
 & \parallel & \parallel \\
 = & Z & \oplus^\perp U \\
 & \text{(cycle)} & \text{(cut)}
 \end{array}$$

The equality is up to the natural identification between X , X' , and E . $\oplus^\perp$ means orthogonal complements.

Theorem 1 resolved Problem 1, giving a complete characterization about the learnable degrees of freedom for a Pauli noise model with general gate sets and noise ansatzes. They also serve as the basis for developing efficient self-consistent learning algorithms. Rigorous statements and detailed proofs for Theorem 1 are presented in Sec. III.

2. Efficient learning algorithms for gate set Pauli noise

We now consider Problem 2, how to efficiently and self-consistently learn a Pauli noise model. Thanks to Theorem 1, to learn a reduced Pauli noise model means to learn every function of the reduced learnable space L_R . It suffices to find a basis for L_R and learn the value of each basis function, as all the other functions can be decided by linearity. This can be done by first finding certain cycle basis for $E = L$ with each basis function learnable from some experiment, and then use the fact $L_R = \mathcal{Q}^\top(L)$ to pick up a subset

of functions to form a basis for L_R . We give such a construction with the additional feature that each experiments uses at most one noisy gate, summarized as follows.

Theorem 2 (Simple experiment construction, informal). There exists a set of $M = \dim(L_R)$ experiments that can determine the value of any function of L_R . Moreover, any experiment from this set either applies one gate from $\mathcal{G}$ exactly once, or applies no gates from $\mathcal{G}$ at all.

To show this, we first show that any cycle containing the root node exactly once (called a *rooted cycle*) can be learned by one experiment. Then notice that, thanks to the connectivity of PTG, there exists a rooted cycle basis where every cycle is of length 2 or 3, corresponding to experiment with no gate or exactly one gate from $\mathcal{G}$. For an efficient noise ansatz, the number of reduced parameters should be at most polynomial in n , which means the number of experiments M needed in Theorem 2 is also polynomial. The protocol can thus be efficiently executed. Details about this protocol are presented in Sec. IV A.

In the literature of quantum noise learning, an important consideration is whether the noise parameter can be learned to relative precision. That is, given a sufficiently small infidelity parameter r of a noisy gate $\tilde{G}$, can we obtain an estimator $\hat{r}$ such that $|\hat{r} - r| \leq O(\epsilon r)$ with high probability, by making a number of measurement that scales as $\epsilon^{-2} \text{polylog}(r^{-1})$? It is known that, if one can concatenate m copies of $\tilde{G}$ to amplify the infidelity to order r^m for any m , this scaling can be achieved [34, Theorem 1]. We should show that such noise amplification is indeed possible for gate set Pauli noise model. We have the following.

Theorem 3 (Amplifying gate noise parameters, informal). For any function that is a cycle vector [see Eq. (15)] and only supported on gate noise parameters, let $\lambda := \exp(-f(x))$. Then, there exists a family of experiments $\{F_m\}_{m \in \mathbb{N}}$ and a Pauli observable, such that the expectation value of the m th experiment is given by $A\lambda^m$, where A is an m -independent parameter.

Theorem 3 ensures one can achieve relative precision for a basis of learnable functions supported on the gate noise parameters, confirming that gate set Pauli noise learning has the same desirable feature as other quantum noise learning schemes such as randomized benchmarking [35]. Details about Theorem 3 are presented in Sec. IV B.

Theorems 2 and 3 address Problem 2 from two different aspects, providing efficient experiment construction for gate set Pauli noise learning. The key is to find an appropriate basis for the reduced learnable space. One caveat is that, since the numbers of edges and nodes in PTG grow exponentially with the number of qubits n , an algorithm

that first finds a cycle basis in PTG and then selects a subset to construct a basis for L_R will inevitably suffer from exponential run time. However, when the noise ansatzes have some structure, we can find such a basis analytically, thus circumventing the run time issue. In what follows, we demonstrate this point by applying our theory to concrete physically relevant noise ansatzes.

3. Applications to spatially-local noise ansatzes

We now apply the general theory of learnability and learning algorithms to concrete noise ansatzes. We will focus on spatially local or quasi-local noise model as an example. Such noise models have been widely studied in the literature of Pauli noise learning [10, 14–17, 20, 22, 36], but our results for the first time give a clear classification of the learnable and gauge degrees of freedom within such models.

We first consider the fully-local model, where the SPAM noise channels are assumed to be qubit-wise product channels, while the gate noise channels act non-trivially only within the gate's support. This model is also sometimes known as a crosstalk-free model and have been considered in, e.g., Refs. [14, 22, 36]. To understand the learnable and gauge spaces for this model, we introduce a family of gauge vectors called the *subsystem depolarizing gauges* (SDGs), denoted by $\{\mathfrak{d}_v\}_{v \in 2^{[n]}}$, where $\mathfrak{d}_v$ represents a gauge transformation as in Eq. (3) with $\mathcal{D}$ chosen as a partially depolarizing channel acting on a subsystem of qubits v . One can show the SDGs form a basis for the complete cut space T . Interestingly, we show that the subset of SDGs acting on one qubit fully characterize the reduced gauge space T_R for a fully-local model.

Theorem 4. For any fully-local noise model $(X_R, \mathcal{Q})$, its reduced gauge space satisfies,

$$\mathcal{Q}(T_R) = \text{span}\{\mathfrak{d}_s : |s| = 1\}. \quad (4)$$

Since $\mathcal{Q}$ is injective, this allows one to uniquely determine T_R . Intuitively, any multi-qubit SDG transformation will violate the locality assumption of qubit-wise independent SPAM noise (and sometimes also local gate noise), and is thus an invalid gauge transformation. We also studied the more involved cases where either SPAM or gates has a fully-local noise assumption while the other does not. Even in that case, we can show that a certain subset of SDGs still span $\mathcal{Q}(T_R)$. Details are presented in Sec. V B.

Regarding learning protocols, the reduced learnable space L_R and a set of $\dim(L_R)$ experiments can be analytically written down. We further study the task of relative precision learning, which boils down to finding a cycle basis for the reduced gate noise parameters. Though we believe this is a hard problem in general, we give a computationally efficient procedure for the case where all Clifford gate from $\mathcal{G}$ acts non-trivially on two qubits, a practical

setting for most existing quantum computing architectures (e.g., Ref. [37]). These issues are addressed in Secs. V C and D.

Going beyond the fully local model, we apply our theory to a quasi-local noise model that describes certain degrees of spatial correlation while remaining parameter-efficient. The quasi-local Pauli noise channels we consider are those whose Pauli eigenvalues factorize according to a factor graph [38]. We first introduce some notations that are needed for defining the quasi-local noise model, defined as follows.

Definition 1. An n -qubit Pauli channel Λ is Ω -local if its Pauli eigenvalues satisfies

$$\lambda_a = \prod_{b \prec a, b \sim \Omega} \exp(-r_b), \quad \forall a \neq I_n, \quad (5)$$

for some $\{r_b \in \mathbb{R} : b \sim \Omega, b \neq I_n\}$ and a factor set $\Omega \subseteq 2^{[n]}$.

Here, $\Omega \subseteq 2^{[n]}$ is a *factor set* if it satisfies $\forall v \in \Omega, \forall \emptyset \subsetneq \mu \subseteq v, \mu \in \Omega$. We write $b \prec a$ if $b_i \neq I \Rightarrow a_i = b_i$ for all $i \in [n]$. We write $b \sim \Omega$ if the support of b belongs to Ω . We show that, under certain ‘‘covariance’’ assumptions of the quasi-local model, the reduced gauge space is again characterized by a subset of SDGs that are consistent with the channel factorization,

Theorem 5 (Informal). For any quasi-local noise model $(X_R, \mathcal{Q})$, given that all Pauli channels are Ω -local and is covariant for all gate in $\mathcal{G}$, then its reduced gauge space satisfies,

$$\mathcal{Q}(T_R) = \text{span}\{\partial_v : v \in \Omega\}. \quad (6)$$

Rigorous statements and detailed proofs are presented in Sec. VI, where we also review useful properties of such a noise model and discuss its connection and distinction with existing models studied in the literature [10, 15–17]. A cycle basis for the reduced learnable space L_R is also analytically constructed for this case.

4. Examples and codes

In Sec. VII, we analyze a few concrete gate sets where the control-Z (CZ) gates are the only multi-qubit entangling gates, with a few spatially-local noise ansatzes. We characterize the learnability of this noise model with more than one noise ansatzes (which are directly related to those that have been studied in the literature [17]) and give efficient learning algorithms in each case. These examples serve as an explicit recipe of gate set Pauli noise learning that is ready to be applied in current quantum devices.

A Python implementation of our gate set Pauli noise learning protocol in the example of parallel CZ on a ring

with nearest-neighbor two-local noise (see Sec. VII C) can be accessed at Ref. [39]. The notebook demonstrates the following pipeline: defining reduced Pauli noise models, designing learning experiments (including relative precision learning and parallelization), verifying the completeness in determining learnable parameters, and a full simulation of the protocol with concrete noise models. Our implementation uses Qiskit [40], an open-source python package for quantum information and quantum computation. We hope this notebook can make it easier for the reader to understand our protocol and adapt it to other gate sets and noise models of their interest.

B. Prior works

Learning the Pauli noise channel Λ affecting a Clifford gate $\mathcal{G}$ (i.e., $\tilde{\mathcal{G}} = \Lambda \circ \mathcal{G}$) in the presence of unknown SPAM noise is a standard task studied in multiple Refs. [10, 12–14, 17, 18, 20–22]. When the ideal Clifford gate $\mathcal{G}$ is equal to the identity, it has been shown that Λ can be fully determined independent of the SPAM noise [10]; When $\mathcal{G}$ is a non-trivial Clifford gate, it is known that certain properties of Λ (e.g., the average fidelity) can be learned independent of SPAM [11], but no existing protocols can completely and SPAM-robustly learn Λ in this case. In fact, it was pointed out [18, 28] that this is a fundamental limitation related to the model non-identifiability, also known as gauge ambiguity in the literature on gate set tomography [27]. Specifically, Ref. [18] have developed a theory that completely characterize the degrees of freedom of a Pauli noise channel associated to any set of Clifford gates and provides a ‘‘gauge-consistent’’ algorithm to learn all the learnable degrees of freedom. The main limitation in Ref. [18] is that it considers the case that each n -qubit gate has a generic n -qubit Pauli noise channel, which has exponentially many parameters, and consequently the learning algorithm would take exponential time. In contrast, efficient parameterization of Pauli noise channels have been studied in the literature and applied in experiments [14–17, 20, 22, 36], but a good understanding about the gauge and learnable degrees of freedom in such models has yet to be obtained, which hinders the application of these protocols (specifically, how to learn such models self-consistently). Our work combines these two lines of research to achieve a self-consistent and efficient Pauli noise learning algorithm.

Our framework resembles averaged circuit eigenvalue sampling (ACES), a Pauli noise learning protocol proposed in Ref. [14] and further analyzed and implemented in Refs. [41] and [22]. Similar to our setting, ACES models every different Clifford gate as having a gate-dependent Pauli noise channel (with an efficient noise ansatz such as fully local noise), and aims at learning every Pauli eigenvalue of every Pauli noise channel. ACES constructs linear equations of the form $\mathbf{b} = A\mathbf{x}$, where $\mathbf{x}$ are the logarithmic

Pauli eigenvalues, $\mathbf{b}$ are the logarithmic circuit eigenvalues which can be experimentally measured, and A is the design matrix determined by a set of Clifford circuits and Pauli measurements. The original ACES effectively fixes a gauge by ignoring state preparation noise, and is thus able to construct a full column-rank A using random Clifford gates. Within ACES’s framework, our work basically says that if SPAM noise parameters are included in $\mathbf{x}$, there is a maximum possible rank of A equal to the number of learnable degrees of freedom, and we provide an explicit recipe for constructing such an A . Moreover, the explicit construction can learn gate parameters to relative precision; see Theorem 9. It is interesting to extend the statistical analysis for ACES [22] to our setting. It is also interesting to study optimal learning and experimental design within our framework [21,42]. Conversely, it is interesting to study the ultimate sample complexity limit for learning gate set Pauli noise models. Such bounds for learning Pauli channels in a black-box setting have been previously investigated [19,43–46].

C. Future directions

1. Quantum error mitigation

One promising practical application of our theory is in quantum error mitigation (QEM) [47]. There have been QEM protocols based on the Pauli noise models [17,48–50] which work by first learning the Pauli noise channels and then either canceling the noise via quasi-probability sampling, or amplifying the noise and then extrapolating to the noiseless limit. Due to the existence of gauge, these protocols generally make assumptions that are not physically justified (e.g., “symmetry assumption” [17,21] or perfect state preparation [50]) in order to determine the noisy Pauli channels. However, it is known that the issue of gauge ambiguity should not hinder self-consistent QEM if assisted by gate set tomography [51,52], although the latter method is not easily scalable. Using our framework, by efficiently learning all the learnable degrees of freedom and anchoring at an arbitrary point of the gauge, one will be able to mitigate the Pauli noise without any additional assumptions. Such a self-consistent and efficient QEM protocol was recently explored [53] after the current paper was posted to arXiv.

2. Beyond Pauli noise model

We believe that our framework and graph-theoretic analysis can be extended beyond the Pauli noise model. We expect the statement that the cycle space equals the learnable space to be held more generally. Note that a Pauli noise model has a linear parameterization (i.e., the log of the Pauli eigenvalues), and one can study the learnable and gauge degrees of freedom as linear subspaces of the parameter spaces. For more general noise models, one might need to study the tangent space of the parameters, which

has been considered in the theory of the first-order gauge-invariant (FOGI) model [54,55]. It is interesting to see if our method can be extended there to have a more quantitative understanding of the learnable and gauge degrees of freedom.

Another important potential extension is to include mid-circuit measurement (MCM, also known as quantum non-demolition measurement), which is a crucial building block for fault-tolerant quantum computation and is being experimentally implemented [56–59]. It has been shown that MCM can be incorporated into Pauli noise models via a generalized randomized compiling technique [60]. More recently, results on the learnability of MCM within the Pauli noise model have been reported [61,62]. It is practically interesting to incorporate those results into the current framework, e.g., to have a self-consistent and efficient protocol for characterizing MCM with a scalable noise ansatz. More generally, it is interesting to study how our framework can be used for noise characterization of early fault-tolerant quantum computing devices [15,16,63].

3. Beyond realizable setting

The learning algorithm we proposed relies on the fact that the noise model of interest is realizable, i.e., there is no out-of-model error. It would be interesting to analyze the effect of such errors. Our current analysis and algorithms require all the noise to be Pauli channels with certain constraints, specifically, the locality constraints as discussed in Secs. V, VI, and VII. In practice, one can fit the same data into models with different locality assumption, and there will be a trade-off between model expressivity and trainability. One can also utilize techniques from structure learning to find the optimal (quasi-)local Pauli noise model (see, e.g., Ref. [64]). It would be interesting to explore learning algorithms with those features within our framework.

II. NOTATIONS AND PRELIMINARIES

This section provides some basic definitions and preliminaries. For more symbols and notations defined later in the paper, see Table I.

A. Bit string and index set

For a positive integer n , define $[n] := \{i \in \mathbb{Z} : 1 \leq i \leq n\}$. We will often use the set of n -bit strings $\mathbb{Z}_2^n$ and the collection of index sets $2^{[n]} := \{\nu : \nu \subseteq [n]\}$. Define the following bijection between $\mathbb{Z}_2^n$ and $2^{[n]}$: $s \in \mathbb{Z}_2^n \leftrightarrow \{j \in [n] : s_j = 1\} \in 2^{[n]}$. Throughout the paper, we often do not distinguish a bit string with an index set, as they can always be identified with each other via the aforementioned bijection. We use $|s|$ to denote the Hamming weight of a bit string s (or its cardinality if understood as an index set). We use 0_n to denote the all-zero bit string (which is $\emptyset$ if

TABLE I. Table of important symbols.

Symbol	Definition or meaning	Reference
X	Parameter space	Definition 3
e_i	Standard basis for parameter space X	Definition 3
$X^{S/M/G}$	Subspace of state prep./meas./gate parameters	Definition 3
L	Learnable space	Lemma 1
T	Gauge space	Lemma 2
E	Edge space of the pattern transfer graph	Definition 8
Z	Cycle space	Definition 9
U	Cut space	Definition 10
X_R	Reduced parameter space	Definition 12
θ_i	Standard basis for reduced parameter space X_R	Definition 12
$\mathcal{Q}$	Embedding map from reduced to complete model	Definition 12
L_R	Reduced learnable space	Definition 13
T_R	Reduced gauge space	Definition 14
L_R^G	Reduced learnable subspace for gate parameters	Definition 16
$\mathfrak{d}_v$	Subsystem depolarizing gauge (SDG)	Definition 19
$\pi^{S/M/G}$	Projection onto subspace of state prep./meas./gate parameters	Proof of Theorem 10
Ω	Factor set of a Markov random field	Sec. VI A
Ω_*	Subset of all maximal factors from Ω	Section VI A
$\Xi_{\mathcal{G}}$	Extended support map	Lemma 10

understood as an index set). We use 1_j to denote the bit string with its j th entry being 1 and other entries being 0 (which is $\{j\}$ if understood as an index set).

We will often use an index set (or, equivalently, a bit string) as the subscript for a string of n objects (e.g., another n -bit string or an n -qubit Pauli operator) to denote a subsequence from the selected indexes. For example, $XYZIZ_{\{1,3\}} = XYZIZ_{10100} = XZ$.

B. Pauli group

We follow the notations of Refs. [10,19]. Let $\mathbf{P}^n = \{I, X, Y, Z\}^{\otimes n}$ be the Pauli group modulo a global phase. $\mathbf{P}^n$ is an Abelian group isomorphic to $\mathbb{Z}_2^{2n}$. Specifically, given a $2n$ -bit string $a = (a_x, a_z) = a_{x,1} \cdots a_{x,n} a_{z,1} \cdots a_{z,n}$, the corresponding Pauli operator is given by

$$P_a = \bigotimes_{j=1}^n i^{a_{x,j} a_{z,j}} X^{a_{x,j}} Z^{a_{z,j}}, \quad (7)$$

where the phase is chosen to ensure Hermiticity. We further define $P_0 := I^{\otimes n} \equiv I_n$ as a shorthand for the n -qubit identity operator.

Besides the standard addition and inner product defined over $\mathbb{Z}_2^{2n}$, we define the symplectic inner product over $\mathbf{P}^n$ as

$$\langle a, b \rangle = \sum_{j=1}^n (a_{x,j} b_{z,j} + a_{z,j} b_{x,j}) \mod 2. \quad (8)$$

One can verify that $\langle a, b \rangle = 0$ if P_a, P_b commute and $\langle a, b \rangle = 1$ otherwise. We sometimes use the label a and the Pauli operator P_a interchangeably when there is no confusion.

Define $\text{pt} : \mathbf{P}^n \mapsto \mathbb{Z}_2^n$ that maps the i th single-qubit Pauli operator to 0 if it is I , or 1 if it is from $\{X, Y, Z\}$. We call $\text{pt}(a)$ the *pattern* of P_a . Alternatively, $\text{pt}(a)$ represents the support of P_a if understood as an index set. The Hamming weight of $\text{pt}(a)$ is called the weight of P_a and is denoted by $w(a)$ or $|a|$.

A k -qubit Pauli P acting on a subset of qubits specified by an index set v is denoted by P_v . Let $W \in \{I, X, Y, Z\}$ be a single-qubit Pauli operator and $v \in 2^{[n]}$, we define W^v to be an n -qubit Pauli such that $(W^v)_j = W$ for $j \in v$ and $(W^v)_j = I$ for $j \notin v$.

C. Pauli channel

An n -qubit Pauli channel Λ has the following two equivalent forms:

$$\Lambda(\rho) = \sum_{a \in \mathbf{P}^n} p_a P_a \rho P_a = \sum_{b \in \mathbf{P}^n} \frac{1}{2^n} \lambda_b P_b \text{Tr}[P_b \rho]. \quad (9)$$

where $\{p_a\}_a$ is called the *Pauli error rates*, while $\{\lambda_b\}_b$ is called the *Pauli eigenvalues* or *Pauli fidelities* (though the latter is potentially misleading since they can in general be negative). These two sets of parameters are related by the Walsh-Hadamard transform [10],

$$\lambda_b = \sum_{a \in \mathbf{P}^n} p_a (-1)^{\langle a, b \rangle}, \quad p_a = \frac{1}{4^n} \sum_{b \in \mathbf{P}^n} \lambda_b (-1)^{\langle a, b \rangle}. \quad (10)$$

Note that for Λ to be a quantum channel, the trace-preserving condition is equivalent to $\sum_a p_a \equiv \lambda_0 = 1$, while the complete positivity is equivalent to $p_a \geq 0, \forall a$. For Λ that has the form of Eq. (9) but does not satisfy these two conditions, we refer to it as a *Pauli diagonal map*.

We will also use the Pauli-Liouville representation. A Hermitian operator O acting on a 2^n -dimensional Hilbert space can be vectorized as a 4^n -dimensional real vector. We denote this vectorization of O as $|O\rangle\rangle$ and the corresponding Hermitian conjugate as $\langle\langle O|$, and define the inner product as $\langle\langle A|B\rangle\rangle := \text{Tr}(A^\dagger B)$, which is the Hilbert-Schmidt inner product in the original space. We also introduce the *normalized Pauli operators* $\{\sigma_a := P_a/\sqrt{2^n}, a \in \mathbb{Z}_2^{2n}\}$, whose vectorization $\{|\sigma_a\rangle\rangle\}_a$ forms an orthonormal basis for the vectorized space. The action of a quantum channel $\mathcal{E}$ becomes matrix multiplication in this representation, i.e., $|\mathcal{E}(\rho)\rangle\rangle = \mathcal{E}^{\text{PL}}|\rho\rangle\rangle$, where $\mathcal{E}^{\text{PL}}$ is the corresponding Pauli-Liouville operator and is usually simply denoted by $\mathcal{E}$ when there is no confusion. In the Pauli-Liouville representation, a Pauli channel is given by

$$\Lambda = \sum_{a \in \mathbb{Z}_2^{2n}} \lambda_a |\sigma_a\rangle\rangle \langle\langle \sigma_a|,$$

which justifies the name of Pauli eigenvalues for $\{\lambda_a\}$.

D. Clifford group

The n -qubit Clifford group $\mathcal{C}^n$ is the group of n -qubit unitary operators that preserve the Pauli group upon conjugation. Mathematically, $\forall C \in \mathcal{C}^n, \forall P_a \in \mathcal{P}^n, CP_a C^\dagger \in \mathcal{P}^n \times \{\pm 1\}$. We often denote the unitary channel of a Clifford gate using the same calligraphic letter, e.g., $\mathcal{C}(\rho) := C\rho C^\dagger$. We also define, with slight abuse of notation, the mapping $\mathcal{C} : \mathbb{Z}_2^{2n} \mapsto \mathbb{Z}_2^{2n}$ via $\mathcal{C}(P_a) \equiv CP_a C^\dagger = \pm P_{\mathcal{C}(a)}$. The Pauli-Liouville operator of a Clifford $\mathcal{C}$ has the following form:

$$\mathcal{C} = \sum_a (\pm 1) |\sigma_{\mathcal{C}(a)}\rangle\rangle \langle\langle \sigma_a|. \quad (11)$$

E. Miscellaneous

For a real linear space X , the dual space of it is denoted by X' . For a linear map Q , its dual map is denoted by $Q^\top$. The image and kernel of a map Q are denoted by $\text{Im}Q$ and $\text{Ker}Q$, respectively. $\mathbb{1}[X]$ is the indicator function that takes 1 if the statement X is true and 0 otherwise. All log represents natural logarithm.

III. LEARNABILITY OF GATE SET PAULI NOISE

A. Basic definitions

In this work, we focus on gate sets consisting of arbitrary single-qubit gates and a set of multi-qubit Clifford gates. We assume time-stationary and Markovian noise that has been twirled into Pauli channels. More formally, we have the following.

Definition 2 (Complete Pauli noise model). Given a set of n -qubit Clifford layers $\mathfrak{G}$, consider the following gate set:

- (1) Initialization: $\rho_0 = |0\rangle\langle 0|^{\otimes n} \mapsto \tilde{\rho}_0 = \Lambda^S(\rho_0)$, where $\Lambda^S = \sum_{a \in \mathcal{P}^n} \lambda_{\text{pt}(a)}^S |\sigma_a\rangle\rangle \langle\langle \sigma_a|$.
- (2) Measurement: $\{E_j = |j\rangle\langle j|\} \mapsto \{\tilde{E}_j = \Lambda^M(|j\rangle\langle j|)\}$, where $\Lambda^M = \sum_{a \in \mathcal{P}^n} \lambda_{\text{pt}(a)}^M |\sigma_a\rangle\rangle \langle\langle \sigma_a|$.
- (3) Single-qubit layer: $\bigotimes_{i=1}^n U_i \mapsto \bigotimes_{i=1}^n U_i$, with all $U_i \in \text{SU}(2)$.
- (4) Multi-qubit Clifford layer: $\mathcal{G} \mapsto \tilde{\mathcal{G}} = \mathcal{G} \circ \Lambda^{\mathcal{G}}$, $\forall \mathcal{G} \in \mathfrak{G}$, where $\Lambda^{\mathcal{G}} = \sum_{a \in \mathcal{P}^n} \lambda_a^{\mathcal{G}} |\sigma_a\rangle\rangle \langle\langle \sigma_a|$.

Furthermore, we assume all $\lambda_{\text{pt}(a)}^S, \lambda_{\text{pt}(a)}^M, \lambda_a^{\mathcal{G}}$ are strictly positive.

In reality, such a model can always be imposed by applying appropriate Pauli twirling sequence, a technique known as randomized compiling [25,26]. The assumption that a single-qubit layer being noiseless can be relaxed to that they have gate-independent noise, in which case such noise can be thought as absorbed into the multi-qubit layer. Note that assuming the SPAM noise channel to be a generalized depolarizing channel (i.e., λ_a only depends on $\text{pt}(a)$) is without loss of generality as they only act on Z basis eigenstates.

Definition 3 (Parameter space). Let $x_a^{S/M/\mathcal{G}} := -\log \lambda_a^{S/M/\mathcal{G}}$ for all $a \neq \mathbf{0}$ for SPAM parameters and all $a \neq I_n$ for all gates $\mathcal{G}$'s parameters. Let $\mathbf{x}$ be a formal vector defined as

$$\mathbf{x} := \sum_{u \neq \mathbf{0}} x_u^S \mathbf{e}_u^S + \sum_{u \neq \mathbf{0}} x_u^M \mathbf{e}_u^M + \sum_{\mathcal{G} \in \mathfrak{G}} \sum_{a \neq I_n} x_a^{\mathcal{G}} \mathbf{e}_a^{\mathcal{G}} \quad (12)$$

where $\{\mathbf{e}_u^S, \mathbf{e}_u^M, \mathbf{e}_a^{\mathcal{G}} : u \neq \mathbf{0}, a \neq I_n, \mathcal{G} \in \mathfrak{G}\}$ forms an orthonormal basis for a real linear space X , called the parameter space. Furthermore, define $X^{S/M} := \text{span}\{\mathbf{e}_u^{S/M} : u \neq \mathbf{0}\}$, $X^{\mathcal{G}} := \text{span}\{\mathbf{e}_a^{\mathcal{G}} : a \neq I_n, \mathcal{G} \in \mathfrak{G}\}$.

Informally, $\mathbf{x}$ is a list of all negative log Pauli eigenvalues in the model, excluding those corresponding to the identity operator as they are always 0 for the noise channels to be trace-preserving. Clearly, any realization of the noise parameters uniquely corresponds to a vector $\mathbf{x}$ in X , but not all $\mathbf{x} \in X$ corresponds to a physical noise model as the channels might not be completely positive.

Definition 4 (Experiment). An experiment is a map $F : X \mapsto \mathbb{R}^{2^n}$ specified by a quantum circuit $\mathcal{C}$ (which is a sequence of finitely many single-qubit layers and multi-qubit Clifford layers within $\mathfrak{G}$), defined as $F(\mathbf{x})_j = \text{Tr}[\tilde{E}_j \tilde{\mathcal{C}}(\tilde{\rho}_0)]$. Here the noisy initial state $\tilde{\rho}_0$, noisy circuit $\tilde{\mathcal{C}}$, and noisy measurement, $\tilde{E}_j$ depends on $\mathbf{x}$.

When $\mathbf{x}$ represents a physical noise model, any experiment $F(\mathbf{x})$ yields a Born's rule probability distribution. However, for mathematical convenience, we naturally extend the domain of F to all over X .

Definition 5 (Learnable function). A real-valued function f on X is **learnable** if there exists a set of experiments $\{F_j\}_{j=1}^m$ such that f is a function of $\{F_j\}_{j=1}^m$. That is, there exists some $\hat{f}$ such that $f(\mathbf{x}) = \hat{f}(F_1(\mathbf{x}), \dots, F_m(\mathbf{x}))$ for all $\mathbf{x} \in X$.

We comment that the learnable function we define here is also known as an identifiable function in the literature of identifiability analysis [65]. Here, for consistency, we will stick to the term “learnable” which has been used in Refs. [18,28].

We will be specifically interested in studying the learnability of linear functions (and we will later explain why that suffices). Note that, there exists an identification between X and its dual space $X' \equiv \mathcal{L}(X, \mathbb{R})$ using the standard inner product, given by $\mathbf{x} \leftrightarrow f_{\mathbf{x}}(\mathbf{y}) := \mathbf{x} \cdot \mathbf{y}$. One can thus think of a linear function on X as a co-vector and denote it with a boldface letter.

Lemma 1. The learnable linear functions form a subspace of X' , called the **learnable space**, denoted by L .

Definition 6 (Gauge vector). A vector $\mathbf{y} \in X$ is a **gauge vector** if for any $\mathbf{x} \in X$, $F(\mathbf{x}) = F(\mathbf{x} + \eta\mathbf{y})$ for any experiment F and $\eta \in \mathbb{R}$.

Lemma 2. The gauge vectors form a linear subspace of X called the **gauge space**, denoted by T .

Proof. By definition, $\mathbf{y} \in X$ being a gauge vector implies $\eta\mathbf{y}$ being a gauge vector for all $\eta \in \mathbb{R}$. Let $\mathbf{y}_1, \mathbf{y}_2 \in X$ be any gauge vectors. By definition, for any $\eta \in \mathbb{R}$ and any observation F ,

$$F(\mathbf{x}) = F(\mathbf{x} + \eta\mathbf{y}_1) = F(\mathbf{x} + \eta\mathbf{y}_1 + \eta\mathbf{y}_2). \quad (13)$$

Therefore, $\mathbf{y}_1 + \mathbf{y}_2$ is also a gauge vector. This proves the gauge vectors forms a linear subspace. ■

Lemma 3. $L \perp T$. That is, for any $\mathbf{f} \in L$ and $\mathbf{y} \in T$ we have $\mathbf{f}(\mathbf{y}) = \mathbf{f} \cdot \mathbf{y} = 0$.

Proof. $\mathbf{y} \in T$ implies $F(\mathbf{x}) = F(\mathbf{x} + \mathbf{y})$ for any experiment F and any $\mathbf{x} \in X$. $\mathbf{f} \in L$ implies there is a set of experiments $\{F_1, \dots, F_m\}$ and a function $\hat{f}$ such that $\mathbf{f}(\mathbf{x}) = \hat{f}(F_1(\mathbf{x}), \dots, F_m(\mathbf{x}))$. Therefore,

$$\begin{aligned} \mathbf{f}(\mathbf{x} + \mathbf{y}) &= \hat{f}(\{F_j(\mathbf{x} + \mathbf{y})\}_{j=1}^m) \\ &= \hat{f}(\{F_j(\mathbf{x})\}_{j=1}^m) = \mathbf{f}(\mathbf{x}) \end{aligned} \quad (14)$$

for any $\mathbf{x}$. Thus $\mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x} + \mathbf{y}) - \mathbf{f}(\mathbf{x}) = 0$ by linearity. This completes the proof. ■

B. Learnability of complete model

In this section, we study the learnability of the complete Pauli noise model. The results presented here are an extension to those in Ref. [18]. However, we will use a quite different proof technique that allows us to generalize the results to noise model with arbitrary constraints in the next section. The following graph is originally defined in Ref. [18], but we twist the definition a bit to include a “SPAM node.”

Definition 7 (Pattern transfer graph with SPAM node (PTG)). The pattern transfer graph associated with an n -qubit Pauli noise model is a directed graph $G = (V, E)$ constructed as follows:

- (1) V : There are $2^n - 1$ nodes v_s for all $s \in \{0, 1\}^n \setminus \{0^n\}$ and 1 additional root node v_R .
- (2) E :
 - (a) For each $a \in \mathcal{P}^n \setminus \{I\}$ and $G \in \mathcal{G}$, construct an edge $v_{\text{pt}(P_a)} \rightarrow v_{\text{pt}(G(P_a))}$ labeled by e_a^G .
 - (b) For each $u \in \{0, 1\}^n \setminus \{0^n\}$, construct an edge $v_R \rightarrow v_u$ labeled by e_u^S .
 - (c) For each $u \in \{0, 1\}^n \setminus \{0^n\}$, construct an edge $v_u \rightarrow v_R$ labeled by e_u^M .

The following definitions from algebraic graph theoretic terminology follow [30,31] with some modifications.

Definition 8. The *edge space* E is a real linear space spanned by a standard basis $\{e\}$ where each e is a formal vector corresponding to an edge $e \in E$.

The reader may notice that $\{e\}$ is used to denote the standard basis both for the parameter space X (Definition 3) and for the edge space E (Definition 8). This is a valid notation as the two spaces are isomorphic. We view the identification of basis elements of the same label as the natural isomorphism between the two spaces.

There are two subspaces of E of special interest, which are defined as follows:

A *walk* is an alternating sequence of vertices and edges $z = (v_{i_0}, e_{i_1}, v_{i_1}, e_{i_2}, v_{i_2}, \dots, v_{i_{q-1}}, e_{i_q}, v_{i_q})$ such that each edge satisfies $e_{i_k} = (v_{i_{k-1}}, v_{i_k})$. We define *cycle* to be a closed walk, which means $v_{i_0} = v_{i_q}$ [66]. We also define a cycle to be *simple* if it does not contain the same edge twice. For a simple cycle z in the PTG, we assign a vector $\mathbf{x}_z \in E$ as follows:

$$\mathbf{x}_z[e] = \begin{cases} +1, & e \in z. \\ 0, & e \notin z. \end{cases} \quad (15)$$

Definition 9. The **cycle space** Z is a linear subspace of E spanned by all simple cycles in G .

A graph is *strongly-connected* if for any two vertices, there exists a cycle containing both of them. The PTG we considered is strongly-connected by construction, as v_R is directed connected to any other nodes.

Given a partition of vertices $V = V_0 \cup V_0^c$, a *cut* is the set of all edges $e = (u, v)$ such that one of u, v belongs to V_0 and the other belongs to V_0^c . For each cut p we assign a vector $y_p \in E$ as follows:

$$y_p(e) = \begin{cases} +1, & e \in p, \quad e \text{ goes from } V_0 \text{ to } V_0^c. \\ -1, & e \in p, \quad e \text{ goes from } V_0^c \text{ to } V_0. \\ 0, & e \notin p. \end{cases} \quad (16)$$

Definition 10. The **cut space** U is a linear subspace of E spanned by all cuts in G .

Below is an important fact about the cycle space Z and cut space U .

Lemma 4 ([31]). $E = Z \oplus^\perp U$. Here $\oplus^\perp$ means an orthogonal direct sum.

The following theorem is a crucial ingredient to understand the learnability of the complete Pauli noise model.

Theorem 6 (Linearity of experiments). Define a cycle to be **rooted** if it contains v_R exactly once.

- (a) For any experiment F , there exists a finite set of linear functions $\{f_j\}_{j=1}^M$ where each f_j corresponds to a rooted cycle, such that F is a function of $\{f_j\}$. That is, there exists a function $\hat{F}$ such that $F(\mathbf{x}) = \hat{F}(f_1(\mathbf{x}), \dots, f_M(\mathbf{x}))$ for all $\mathbf{x} \in X$.
- (b) For any linear function f_j corresponding to a rooted cycle, there exists an experiment F such that f_j is a function of F . That is, there exists a function $\hat{f}_j$ such that $f_j(\mathbf{x}) = \hat{f}_j(F(\mathbf{x}))$ for all $\mathbf{x} \in X$.

This is detailed in Ref. [18]. For completeness, we give a proof in Appendix A. Intuitively, Theorem 6(b) says that any rooted cycle can be learned from at least one experiment. In fact, the experiment involves preparing a certain noisy Pauli eigenstate, applying a sequence of perfect single-qubit and noisy multi-qubit Clifford gates, and measuring a certain noisy Pauli observable. The expectation value will have the form of $\lambda_{b_0}^S \lambda_{p_1}^{G_1} \cdots \lambda_{p_m}^{G_m} \lambda_{b_m}^M$, taking a log of which yields the desired rooted cycle; Theorem 6(a) says that rooted cycles are complete, in the sense that any experiment (possibly involving non-stabilizer states, measurements, and non-Clifford single-qubit gates) can be predicted using linear combinations of rooted cycles. Therefore, it is sufficient to characterize the learnability of linear functions and not to worry about the non-linear ones.

To understand the full picture of learnable and gauge degrees of freedom, we introduce the following definition.

Definition 11 (Observation map). Fix an arbitrary basis $\{f_i\}_{i=1}^{|Z|}$ for Z such that each f_i represents a rooted cycle. Define the **observation map** $\mathcal{P} : X \mapsto \mathbb{R}^{|Z|}$ via

$$\mathcal{P}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_{|Z|}(\mathbf{x})]^T. \quad (17)$$

Note that, thanks to the strong connectivity of the PTG, a cycle basis consisting of only rooted cycles always exist. One construction is given below.

Lemma 5 (A rooted cycle basis). The following set of vectors forms a basis for Z :

$$\mathbf{B} := \{e_u^S + e_u^M : u \neq 0_n\} \cup \{e_{\text{pt}(a)}^S + e_a^G + e_{\text{pt}(\mathcal{G}(a))}^M : \mathcal{G} \in \mathfrak{G}, a \neq I_n\}. \quad (18)$$

Proof. To begin with, every vector in $\mathbf{B}$ is obviously a rooted cycle. Since the PTG is strongly connected by construction, the dimension of T is $2^n - 1$ [31], thus

$$|Z| = |E| - |T| = 2^n - 1 + |\mathfrak{G}|(4^n - 1), \quad (19)$$

which is the same as the cardinality of $\mathbf{B}$. Therefore, it suffices to show the linear independence for $\mathbf{B}$. Consider the equation $\sum_{i: e_i \in \mathbf{B}} \alpha_i e_i = \mathbf{0}$. Note that each e_a^G appears in exactly one vector from the second part of $\mathbf{B}$, for which the corresponding α_i must be all zero, but the first part is also obviously linearly independent, which means all α_i must be zero, proving the linear independence. ■

Theorem 7. Up to identification between X and X' using the standard basis,

$$\begin{array}{ccc} \text{Im } \mathcal{P}^\top & \text{Ker } \mathcal{P} & \\ \parallel & \parallel & \\ X = L \oplus^\perp T & & \\ \parallel & \parallel & \parallel \\ E = Z \oplus^\perp U & & \end{array} \quad (20)$$

In words, the cycle space Z is equal to the learnable space L , while the cut space U is equal to the gauge space T . Besides, the two spaces are orthogonal complements to each other. Finally, the learnable space L is also equal to the image for the dual of the observation map $\mathcal{P}$, while the gauge space T is the kernel of $\mathcal{P}$.

Remark. While $E = Z \oplus^\perp U$ is a graph-theoretic fact, we do not know *a priori* whether $X = L \oplus T$ holds under our definition.

Proof. We first show that $T = \text{Ker}(\mathcal{P})$, i.e., $\mathbf{y} \in T \Leftrightarrow \mathcal{P}(\mathbf{y}) = \mathbf{0}$. Indeed, Theorem 6(a) implies that

$F(\mathbf{x}) = \widehat{F}(\mathcal{P}(\mathbf{x}))$ for some function $\widehat{F}$. Thus, for any $\mathbf{y} \in \text{Ker}(\mathcal{P})$ and $\eta \in \mathbb{R}$,

$$F(\mathbf{x} + \eta\mathbf{y}) = \widehat{F}(\mathcal{P}(\mathbf{x} + \eta\mathbf{y})) = \widehat{F}(\mathcal{P}(\mathbf{x})) = F(\mathbf{x}). \quad (21)$$

The second equality uses the linearity of $\mathcal{P}$. Therefore, $\text{Ker}(\mathcal{P}) \subseteq T$. On the other hand, for any $\mathbf{y} \notin \text{Ker}(\mathcal{P})$, there is at least one f_i such that $f_i(\mathbf{y}) \neq 0$. Now, Theorem 6(b) implies that there exists an experiment F_i and a function $\hat{f}_i$ such that $f_i(\mathbf{x}) = \hat{f}_i(F_i(\mathbf{x}))$ for all $\mathbf{x}$. Thus,

$$\begin{aligned} f_i(\mathbf{y}) \neq 0 &\Rightarrow f_i(\mathbf{x}) \neq f_i(\mathbf{x} + \mathbf{y}) \\ &\Rightarrow \hat{f}_i(F(\mathbf{x})) \neq \hat{f}_i(F(\mathbf{x} + \mathbf{y})) \\ &\Rightarrow F(\mathbf{x}) \neq F(\mathbf{x} + \mathbf{y}) \\ &\Rightarrow \mathbf{y} \notin T. \end{aligned} \quad (22)$$

Therefore, $T \subseteq \text{Ker}\mathcal{P}$. This proves the claim that $T = \text{Ker}\mathcal{P}$.

We then show that $L = \text{Im}(\mathcal{P}^\top)$. For any $\mathbf{f} \in \text{Im}(\mathcal{P}^\top)$, there is some $\mathbf{h} \in \mathcal{L}(\mathbb{R}^{|Z|}, \mathbb{R})$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathcal{P}(\mathbf{x}))$, $\forall \mathbf{x} \in X$. Theorem 6(b) implies that each entry of $\mathcal{P}(\mathbf{x})$ is learnable, thus $\mathbf{f}$ is learnable. This means $\text{Im}(\mathcal{P}^\top) \subseteq L$; On the other hand, $L \perp T = \text{Ker}(\mathcal{P}) \Rightarrow L \subseteq \text{Im}(\mathcal{P}^\top)$ as the kernel is the orthogonal complement of the adjoint image. This proves the claim that $L = \text{Im}(\mathcal{P}^\top)$.

Finally, it is not hard to see that $\text{Ker}(\mathcal{P}) = U$ thanks to the duality between cycle space and cut space [31], and hence $\text{Im}(\mathcal{P}^\top) = Z$. This completes the proof of Theorem 7. ■

C. Learnability of the reduced model

Since the complete noise model contains exponentially many parameters, we often need to make additional assumptions in practice to reduce its complexity. Here, we want to understand the learnability of such a reduced noise model.

Definition 12. A reduced Pauli noise model is defined by a tuple $(X_R, \mathcal{Q})$, where

- (1) X_R is a real linear space called the **reduced parameter space**. A realization of the reduced noise parameters $\{r_i\}_{i=1}^{|X_R|}$ corresponds to a vector $\mathbf{r} \in X_R$ via $\mathbf{r} = \sum_i r_i \boldsymbol{\vartheta}_i$ where $\{\boldsymbol{\vartheta}_i\}$ is a standard basis for X_R .
- (2) $\mathcal{Q} : X_R \mapsto X$ is an injective linear map called the **embedding map** such that for any $\mathbf{r}$, the corresponding realization of the complete noise parameters is given by $\mathbf{x} := \mathcal{Q}(\mathbf{r})$.

Note that we require $\mathcal{Q}$ to be linear and injective. We will see in the following sections that many physically-motivated reduced noise models (e.g., spatially local

noise) can indeed be described by a linear embedding map. The requirement of injectivity, equivalent to $\text{Ker}(\mathcal{Q}) = \{\mathbf{0}\}$, is for mathematical convenience. In practice, if the embedding map is not injective, one can always construct another reduced model with fewer parameters and admits an injective embedding map.

Any experiment F as in Definition 4 induces a function $F \circ \mathcal{Q}$ acting on X_R . The definitions and properties of learnable functions and gauge vectors then carry over to the reduced model, by replacing X by X_R in Definitions 5 and 6. Formally, we have the following.

Definition 13 (Reduced learnable space). A real-valued function f on X_R is **learnable** if there exists a finite set of experiments $\{F_j\}_{j=1}^m$ such that f is a function of $\{F_j \circ \mathcal{Q}\}_{j=1}^m$. That is, there exists some $\hat{f}$ such that $f(\mathbf{r}) = \hat{f}(F_1(\mathcal{Q}(\mathbf{r})), \dots, F_m(\mathcal{Q}(\mathbf{r})))$ for all $\mathbf{r} \in X_R$. The linear subspace of X'_R spanned by all linear learnable functions is called the **reduced learnable space**, denoted by L_R .

Definition 14 (Reduced gauge space). A vector $\mathbf{y} \in X_R$ is a **gauge** vector if for any $\mathbf{r} \in X_R$, $F(\mathcal{Q}(\mathbf{r})) = F(\mathcal{Q}(\mathbf{r} + \eta\mathbf{y}))$ for any experiment F and $\eta \in \mathbb{R}$. The linear subspace of X_R spanned by all the gauge vectors is called the **reduced gauge space**, denoted by T_R .

The characterization of learnability naturally extends to the reduced model, as stated in the next theorem.

Theorem 8. $L_R = \text{Im}(\mathcal{Q}^\top \circ \mathcal{P}^\top)$, $T_R = \text{Ker}(\mathcal{P} \circ \mathcal{Q})$.

We provide a proof for Theorem 8 in Sec. B. Basically, all steps in the proof of Theorem 7 naturally carry over. The following corollary characterizes the relation of the learnable and gauge spaces between the complete and the reduced noise model.

Corollary 1. $L_R = \mathcal{Q}^\top(L)$, $T_R = \mathcal{Q}^{-1}(T)$.

Here $\mathcal{Q}^{-1}(T) := \{\mathbf{x} \in X_R \mid \mathcal{Q}(\mathbf{x}) \in T\}$ is the preimage of T under $\mathcal{Q}$. Alternatively, $T_R = \mathcal{Q}^{-1}(T \cap \text{Im}(\mathcal{Q}))$ where $\mathcal{Q}^{-1}$ is viewed as the inverse of $\mathcal{Q}$ restricted to its image (recall that $\mathcal{Q}$ is assumed to be injective).

Proof. For any $\mathbf{f} \in X'_R$, $\mathbf{f} \in L_R \Leftrightarrow \exists \mathbf{g} \in \mathcal{L}(\mathbb{R}^{|Z|}, \mathbb{R})$ s.t. $\mathbf{f} = \mathcal{Q}^\top(\mathcal{P}^\top(\mathbf{g})) \Leftrightarrow \mathbf{f} \in \mathcal{Q}^\top(L)$. Thus, $L_R = \mathcal{Q}^\top(L)$.

Let us consider $\mathcal{Q}(T_R)$. Theorems 7, 8 imply that $\mathcal{Q}(T_R) \subseteq T$. Also, $\mathcal{Q}(T_R) \subseteq \text{Im}(\mathcal{Q})$ trivially holds. On the other hand, for any $\mathbf{y} \in T \cap \text{Im}(\mathcal{Q})$, $\mathcal{P}(\mathbf{y}) = 0$, and there exists some $\mathbf{x} \in X_R$ such that $\mathbf{y} = \mathcal{Q}(\mathbf{x})$, but this implies $\mathcal{P}(\mathcal{Q}(\mathbf{x})) = 0$ and thus $\mathbf{x} \in T_R$, $\mathbf{y} \in \mathcal{Q}(T_R)$. We conclude that $\mathcal{Q}(T_R) = T \cap \text{Im}(\mathcal{Q})$. Finally, note that $\mathcal{Q}$ is injective by assumption and is thus invertible when restricted to its image. We conclude that $T_R = \mathcal{Q}^{-1}(T \cap \text{Im}(\mathcal{Q}))$. ■

IV. ALGORITHMS FOR LEARNING GATE SET PAULI NOISE

We have established a theory for characterizing the learnable and gauge degrees of freedom for any complete or reduced Pauli noise models. The next question to be asked is how to design experiments to learn the learnable degrees of freedom, and how efficiently one can learn them. In this section, we first introduce a general algorithm for learning Pauli noise models, which leads to efficient experimental designs. We then discuss the issue of learning noise parameters to additive and multiplicative precision. The algorithms discussed here will be applied in the next section to address many practically interesting examples of Pauli noise models.

From this section, we focus on Clifford circuits as they are sufficient for learning Pauli noise. Also, instead of describing an experiment with its measurement probability distribution, we will talk about expectation values of measuring certain Pauli observable. Since we assume noiseless single-qubit layers, measuring Pauli P_a effectively picks up $\lambda_{\text{pt}(a)}^M$ from the measurement noise channel. We thus define the effective noisy observable as $\tilde{P}_a := \lambda_{\text{pt}(a)}^M P_a$. Similarly, we use ρ_a to denote the +1 eigenstate of P_a obtained by applying an appropriate single-qubit Clifford layer to $\rho_0 \equiv |0\rangle\langle 0|^{\otimes n}$ [67]. The noisy realization of ρ_a is denoted as $\tilde{\rho}_a$ and satisfies $\text{Tr}[P_a \tilde{\rho}_a] = \lambda_{\text{pt}(a)}^S$. Finally, we use the tuple $(\rho_a, \mathcal{C}, P_b)$ to denote an experiment that (ideally) starts with ρ_a , applies the sequence of gates $\mathcal{C}$, and measures the observable P_b .

A. Basic definitions and algorithm

Let us first define the task of learning a Pauli noise model. Given a reduced noise model described by an embedding map $\mathcal{Q}$. Recall that L_R describes the linear space for all learnable functions about the noise model (and that Theorem 6 says *linear* learnable functions suffice). Thus, we just need to do the following:

- (1) Specify a basis for L_R , denoted as $\{\mathbf{f}_j\}_{j=1}^{|L_R|}$,
- (2) Specify the experiments to learn each $\mathbf{f}_j$.

By conducting those experiments, we will have effectively learned every function in L_R , which are all we can learn about the noise model.

However, for an arbitrary basis vector $\mathbf{f}_j \in L_R$, it is not straightforward to find the experiments that learn it, despite their existence. We would like to find a reduced basis that is more closely related to experiments. Recall that for the complete model, the learnable space L corresponds to the cycle space Z of PTG, and each rooted cycle can be extracted from one experiment [Theorem 6(b)]. This motivates the following definition.

ALGORITHM 1. A simple algorithm for learning gate set Pauli noise.

-
- (1) Select a subset $\mathbf{B}_R \subseteq \mathbf{B}$ such that $\mathcal{Q}^\top(\mathbf{B}_R)$ forms a basis for L_R .
 - (2) Learn all $\mathbf{e}_u^S + \mathbf{e}_u^M \in \mathbf{B}_R$ by preparing $\tilde{\rho}_0$ and measuring $\tilde{Z}^u$.
 - (3) Learn all $\mathbf{e}_{\text{pt}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{pt}(\mathcal{G}(a))}^M \in \mathbf{B}_R$ by preparing $\tilde{\rho}_a$, applying $\tilde{\mathcal{G}}$, and measuring $\tilde{P}_{\mathcal{G}(a)}$.
-

Definition 15 (Reduced cycle). For any $\mathbf{f} \in L_R$, if there exists a $\mathbf{g} \in L$ that is a cycle and that $\mathbf{f} = \mathcal{Q}^\top(\mathbf{g})$, we call $\mathbf{f}$ a **reduced cycle** corresponding to $\mathbf{g}$. If $\mathbf{g}$ is also a rooted cycle, we call $\mathbf{f}$ a **rooted reduced cycle**.

We remind the reader that not all $\mathbf{g} \in L$ represents a cycle, as some might only be linear combinations of cycles. Since Corollary 1 states that $L_R = \mathcal{Q}^\top(L)$, given a rooted cycle basis $\{\mathbf{g}_i\}_{i=1}^{|L|}$ for L , we know that $\{\mathcal{Q}^\top(\mathbf{g}_i)\}_{i=1}^{|L|}$ spans L_R , thus we can pick a subset $\{\mathbf{f}_j\}_{j=1}^{|L_R|} \subseteq \{\mathcal{Q}^\top(\mathbf{g}_i)\}_i$ as a basis for L_R consisting solely of rooted reduced cycles. Theorem 6(b) thus immediately gives an experiment to learn $\mathbf{f}_j$.

As a concrete example, consider the rooted cycle basis introduced in Lemma 5,

$$\mathbf{B} := \{\mathbf{e}_u^S + \mathbf{e}_u^M : u \neq 0\} \cup \{\mathbf{e}_{\text{pt}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{pt}(\mathcal{G}(a))}^M : \mathcal{G} \in \mathfrak{G}, a \neq I\}. \quad (23)$$

We have the following procedure for noise learning.

As an obvious challenge to this approach, the cardinality of $\mathbf{B}$ is $2^n - 1 + |\mathfrak{G}|(4^n - 1)$, scaling exponentially with n . Even if the dimension of L_R is polynomial in n , step 1 of picking a basis for L_R still requires exponential computational complexity in general. However, for physically-motivated noise ansatzes such as spatially local noise, it is often easy to efficiently construct such a basis. This will be discussed in details when we turn to concrete noise models. Conditioned on that $\mathbf{B}_R$ can be efficiently constructed, steps 2 and 3 involve at most $|L_R|$ experiments to run, which is polynomial in n for any feasible reduced noise model. One might notice that this algorithm resembles standard quantum process tomography [68]. Specifically, each gate from the gate set is learned one by one. No concatenation or composition of different gates is used.

B. Learning with relative precision

In practice, there are often considerations beyond merely constructing a complete set of experiments. One important factor to consider is the precision-efficiency trade-off in estimating the noise parameters. Many existing RB-type noise characterization protocols can learn the gate noise parameter to relative precision (with respect to the infidelity) using a small number of initialization

and measurements [10, 11, 34]. More precisely, suppose we are interested in learning certain gate fidelity parameter λ , RB-type protocol can generally construct a family of experiments $\{(\rho_0, \mathcal{C}^d, P)\}_d$ such that the circuit fidelities satisfy

$$f_d := \text{Tr}[\tilde{P}\tilde{\mathcal{C}}^d(\tilde{\rho}_0)] = \alpha\lambda^d, \quad (24)$$

where α represents some SPAM fidelity, d is the number of concatenation of the relevant gates. We assume both α and λ are sufficiently close to 1. Now choose $d \in \{0, m\}$, estimate each $\hat{f}_d := f_d + \hat{\varepsilon}_m$, and construct the following ratio estimator:

$$\begin{aligned} \hat{\lambda} &:= \frac{\hat{f}_m}{\hat{f}_0} = \left(\frac{\alpha\lambda^m + \hat{\varepsilon}_m}{\alpha + \hat{\varepsilon}_0} \right)^{\frac{1}{m}} \\ &\approx \lambda \left(1 + \frac{\hat{\varepsilon}_m\lambda^{-m} - \hat{\varepsilon}_0}{m\alpha} \right). \end{aligned} \quad (25)$$

The last approximation holds given that $|\hat{\varepsilon}_0/\alpha|, |(\hat{\varepsilon}_m\lambda^{-m} - \hat{\varepsilon}_0)/m\alpha| \ll 1$. If we choose $1/m \approx \log \lambda^{-1} = 1 - \lambda + o(1 - \lambda)$, we would obtain that

$$\hat{\lambda} \approx \lambda \left(1 + (1 - \lambda) \frac{\hat{\varepsilon}_m e - \hat{\varepsilon}_0}{\alpha} \right). \quad (26)$$

This means as long as we estimate f_0, f_m to $\pm \varepsilon$ additive precision, we would obtain $O(\varepsilon)$ relative precision estimation for λ with respect to the infidelity $1 - \lambda$. As a side note, one can efficiently find the required m using a search subroutine described in Ref. [34, Sec. V]. The basic idea is to estimate f_d for d 's chosen from an exponentially increasing list, and output $m = d$ when $\hat{f}_d/\hat{f}_0$ is below certain threshold.

The above analysis shows that, as long as one can construct a family of experiments as in Eq. (24), one can amplify λ by concatenation and learn it to relative precision. We now explain how to fit this into the framework of gate set Pauli noise learning.

Definition 16. For any reduced Pauli noise model $(X_R, \mathcal{Q})$, the reduced learnable space for gate parameters is defined as $L_R^G := \mathcal{Q}^T(L \cap X^G)$.

Theorem 9 (Noise parameter amplification). Given a reduced model $(X_R, \mathcal{Q})$, for any reduced cycle $\mathbf{f} \in L_R^G$, there exists a family of experiments $\{(P, \mathcal{C}^m, \rho_P)\}_m$ and a rooted cycle $\alpha \in L$ such that for all $m \in \mathbb{N}$,

$$\text{Tr}[\tilde{P}\tilde{\mathcal{C}}^m(\tilde{\rho}_P)] = \exp(-\alpha(\mathcal{Q}(\mathbf{r}))) \exp(-m\mathbf{f}(\mathbf{r})). \quad (27)$$

In other words, any functions on the reduced model that correspond to learnable functions consisting solely of gate noise parameters in the complete model can be amplified via concatenation, and can thus be learned to relative precision. Here, the gate sequence $\mathcal{C}$ can be a composition

ALGORITHM 2. Relative precision learning of gate set Pauli noise.

-
- (1) Select a reduced cycle basis of L_R^G , denoted as $\mathbf{B}_R^G$.
 - (2) Augment $\mathbf{B}_R^G$ into a basis for L_R using vectors from $\mathcal{Q}^T(\mathbf{B})$.
 - (3) Learn all vectors from $\mathbf{B}_R^G$ by running the corresponding set of experiments $\{(P, \mathcal{C}^m, \rho_P)\}_m$ described by Theorem 9.
 - (4) Learn all the other basis vectors using steps 2 and 3 from Algorithm 1.
-

of multiple gates from the gate set interleaved by noiseless single-qubit gates, which is similar to the concept of “germs” in gate set tomography [27], and also appears in other Pauli noise learning protocols [17, 18, 20].

Proof. By definition, there exists a cycle $\mathbf{g} \in L \cap X^G$ such that $\mathbf{f} = \mathcal{Q}^T(\mathbf{g})$. Let u be any vertex of the PTG that is contained in $\mathbf{g}$. Then $\alpha := \mathbf{e}_u^S + \mathbf{e}_u^M$ is a rooted cycle also containing u , and it is not hard to see that $\alpha + m\mathbf{g}$ for any $m \in \mathbb{N}$ forms a rooted cycle. Theorem 6(b) then implies that there exists an experiment $(P, \mathcal{C}_m, \rho_P)$ such that

$$(\alpha + m\mathbf{g}) \cdot \mathcal{Q}(\mathbf{r}) = -\log(\text{Tr}[\tilde{P}\tilde{\mathcal{C}}_m(\tilde{\rho}_P)]). \quad (28)$$

Finally, it is clear from the proof of Theorem 6(b) that each $\mathcal{C}_m$ can actually be chosen as $\mathcal{C}^m$ (i.e., concatenating $\mathcal{C}$ m times) for an appropriate gate sequence $\mathcal{C}$. This completes the proof. ■

In light of Theorem 9, one can design an algorithm that not only learns everything in L_R , but also learn a basis of L_R^G to relative precision.

Again, the major challenge is to efficiently construct the reduced cycle basis, which would generally require doing linear algebra on an exponential-dimensional linear space. We will soon see that for some physically-motivated noise ansatz and realistic gate set, such a basis can be constructed efficiently.

We close this section with one final remark: Although specifying a reduced cycle basis of L_R^G is sufficient to specify any functions therein, learning each basis function to relative precision does not guarantee relative precision estimation for arbitrary functions. For example, relative-precision estimators for both $\mathbf{f}_1$ and $\mathbf{f}_2$ does not guarantee a relative-precision estimator for $\mathbf{f}_1 - \mathbf{f}_2$, especially when the true value of $\mathbf{f}_1$ and $\mathbf{f}_2$ are close to each other. In such cases, one might need to conduct more experiments than merely learning a basis in order to retrieve relative precision estimation for those functions. We will leave this issue for future study.

V. APPLICATIONS TO FULLY LOCAL NOISE MODEL

A. Basic definitions

We have established general results on the learnability of any reduced Pauli noise models. Now we apply our theory to concrete physically-motivated noise ansatz. In this section, we will focus on the fully local noise model, or the so-called crosstalk-free noise model. As we will soon see, the reduced gauge space in this case is generated by a certain subsystem depolarizing channel. We will also give efficient construction of the learning algorithms.

Before defining into the spatially local Pauli noise model, let us first introduce the *layer-uncorrelated noise model*.

Definition 17. For a reduced noise model $(X_R, \mathcal{Q})$, if X_R and $\mathcal{Q}$ can be decomposed as

$$\begin{aligned} X_R &= X_R^S \oplus X_R^M \bigoplus_{\mathcal{G} \in \mathfrak{G}} X_R^{\mathcal{G}}, \\ \mathcal{Q} &= \mathcal{Q}^S \oplus \mathcal{Q}^M \bigoplus_{\mathcal{G} \in \mathfrak{G}} \mathcal{Q}^{\mathcal{G}}, \end{aligned} \quad (29)$$

where $\mathcal{Q}^{\square} : X_R^{\square} \mapsto X^{\square}$ satisfies $\mathcal{Q}^{\square}(r^{\square}) = x^{\square}$ for all $\square \in \{S, M\} \cup \mathfrak{G}$, then we call it a **layer-uncorrelated noise model**.

In words, the noise channel for different layers are determined by disjoint sets of reduced noise parameters. The noise parameters for one layer has no effect on any other layer. Many noise models studied in the literature of quantum benchmarking satisfies this assumption (e.g., Refs. [14, 20]). We note that this is *in contrast* to the well-known “gate-independent noise model” [69–73] where all layers of gates have the same noise channel. Though we will not consider the latter in this section, the general framework established in the previous section is still applicable.

Lemma 6. For a layer-uncorrelated model $(X_R, \mathcal{Q})$, the reduced gauge space T_R satisfies

$$\mathcal{Q}(T_R) = \bigcap_{\square \in \{S, M\} \cup \mathfrak{G}} \left\{ \mathbf{z} \in T : \mathbf{z}^{\square} \in \text{Im} \mathcal{Q}^{\square} \right\}. \quad (30)$$

Proof. Corollary 1 says $\mathcal{Q}(T_R) = T \cap \text{Im} \mathcal{Q}$, which satisfies

$$\begin{aligned} T \cap \text{Im} \mathcal{Q} &= \{ \mathbf{z} \in T : \mathbf{z} \in \text{Im} \mathcal{Q}^S \oplus \text{Im} \mathcal{Q}^M \bigoplus_{\mathcal{G} \in \mathfrak{G}} \text{Im} \mathcal{Q}^{\mathcal{G}} \} \end{aligned}$$

$$\begin{aligned} &= \{ \mathbf{z} \in T : \mathbf{z}^{\square} \in \text{Im} \mathcal{Q}^{\square}, \forall \square \in \{S, M\} \cup \mathfrak{G} \} \\ &= \bigcap_{\square \in \{S, M\} \cup \mathfrak{G}} \left\{ \mathbf{z} \in T : \mathbf{z}^{\square} \in \text{Im} \mathcal{Q}^{\square} \right\}. \end{aligned} \quad (31)$$

■

As a result, we can separately analyze the gauge vectors allowed by state preparation, measurements, and each layer of gates. The intersection of all of them will give the correct reduced gauge space. In what follows, we will use this approach to analyze the gauge space for the fully local noise model, and for the quasi-local noise model in the next section.

B. Learnability of fully local model

Let us first define the fully local noise model. For any $\mathcal{G} \in \mathfrak{G}$, we use $\text{supp}(\mathcal{G})$ to denote the support of $\mathcal{G}$, i.e., the subsystem of qubits that $\mathcal{G}$ non-trivially acts on. The restriction of $\mathcal{G}$ to its support is denoted by $\hat{\mathcal{G}}$. The number of qubits in the support is denoted by $|\mathcal{G}|$. Recall that we use $\{\mathbf{r}_i\}$ and $\{\mathbf{e}_i\}$ to denote the standard basis for X_R and X , respectively. That is, the reduced and complete noise parameters are encoded into vectors according to $\mathbf{r} = \sum_i r_i \mathbf{r}_i$ and $\mathbf{x} = \sum_i x_i \mathbf{e}_i$, respectively.

We will first give the formal definition of the fully local noise model, and explain the intuition in a moment. [Recall that, for an n -qubit Pauli b , we use $b_{\text{supp}(\mathcal{G})}$ to denote the substring of Pauli operators within $\text{supp}(\mathcal{G})$.]

Definition 18 (Fully local noise). Consider a layer-uncorrelated noise model $(X_R, \mathcal{Q})$:

- (1) If the state preparation noise parameters are given by $\{r_j^S : j \in [n]\}$ and it holds that

$$\mathcal{Q}(\mathbf{r}_j^S) = \sum_{u \neq 0_n} \mathbb{1}[u_j = 1] \mathbf{e}_u^S, \quad \forall j \in [n], \quad (32)$$

we say the model has **fully local state preparation noise**. Similar for the measurement.

- (2) If the noise parameters for some $\mathcal{G} \in \mathfrak{G}$ are given by $\{r_a^{\mathcal{G}} : a \in \mathbf{P}^{|\mathcal{G}|}, a \neq I_{|\mathcal{G}|}\}$ and it holds that

$$\begin{aligned} \mathcal{Q}(\mathbf{r}_a^{\mathcal{G}}) &= \sum_{b \neq I_n} \mathbb{1}[b_{\text{supp}(\mathcal{G})} = a] \mathbf{e}_b^{\mathcal{G}}, \\ &\quad \forall a \in \mathbf{P}^{|\mathcal{G}|}, a \neq I_{|\mathcal{G}|}, \end{aligned} \quad (33)$$

we say the model satisfies fully local noise for $\mathcal{G}$. If this holds for all $\mathcal{G} \in \mathfrak{G}$ we say the model has **fully local gate noise**.

The model is **fully local** if it has fully local state preparation, measurement, and gate noise.

The following lemma describes the transformation between the reduced parameters and the complete parameters within a fully local model, as an immediate consequence of the above definitions.

Lemma 7. If $(X_R, \mathcal{Q})$ satisfies fully local state preparation noise, then

$$\begin{cases} x_u^S = \sum_{j \in [n]} \mathbb{1}[u_j = 1] r_j^S, & \forall u \in \{0, 1\}^n, u \neq 0_n. \\ r_j^S = x_{1_j}^S, & \forall j \in [n]. \end{cases} \quad (34)$$

Similar for the measurement side; If $(X_R, \mathcal{Q})$ satisfies fully local noise for some gate $\mathcal{G}$, then

$$\begin{cases} x_b^G = r_{b_{\text{supp}(\mathcal{G})}}^G, & \forall b \in \mathbf{P}^n, b \neq I_n. \\ r_a^G = x_b^G \text{ s.t. } b_{\text{supp}(\mathcal{G})} = a, & \forall a \in \mathbf{P}^{|\mathcal{G}|}, a \neq I_{|\mathcal{G}|}. \end{cases} \quad (35)$$

Proof. By definition of the fully local noise model, one can easily verify that

$$\mathbf{x}^S = \mathcal{Q}^S(\mathbf{r}^S) = \sum_{u \neq 0_n} \sum_{j \in [n]} \mathbb{1}[u_j = 1] r_j^S \mathbf{e}_u^S. \quad (36)$$

$$\mathbf{x}^G = \mathcal{Q}^G(\mathbf{r}^G) = \sum_{b \neq I_n} r_{b_{\text{supp}(\mathcal{G})}}^G \mathbf{e}_b^G. \quad (37)$$

Comparing these with the definitions of $\mathbf{x}$ and $\mathbf{r}$ immediately gives us the transformation from $\{r_i\}$ to $\{x_i\}$. The correctness of the inverse transformation can be verified by substitution. ■

In words, the fully local state preparation or measurement noise is a qubit-wise independent channel, and the fully local gate noise is a channel acting only on the gate's support and inducing no crosstalk on idling qubits. See Fig. 3(a) for an illustration. We study these cases separately, as there could be situations where the SPAM noise is local while gate noise is not, or the other way around.

We also note that whether fully local gate noise describes contextual noise depends on the definition of $\mathfrak{G}$: If one define $\mathfrak{G}$ in a way that each $\mathcal{G} \in \mathfrak{G}$ acts non-trivially at, say, at most 2-qubit. Let $\mathcal{G}_1, \mathcal{G}_2$ be two Clifford layers from $\mathfrak{G}$ that act non-trivially on disjoint sets of qubits. The noise channel for applying both gates in parallel can then be viewed as the tensor product of two associated noise channels. However, if one believe the parallel application should suffer from a different noise channel, one can add it as an independent layer into $\mathfrak{G}$. Although one needs to be careful not to make the size $\mathfrak{G}$ exponentially large.

The advantages of a fully-local noise model is a small number of model parameters [as small as $O(n)$ for a system with a linear connectivity], which makes working with such model potentially computationally efficient. Previous work, including ACES [14], has focused on characterizing such a noise model. It is thus of great interest to understand the learnable and gauge degrees of freedom within this model. For this purpose, let us introduce the following set of gauge vectors.

Definition 19 (Subsystem depolarizing gauge, SDG). For any $s \in \{0, 1\}^n \setminus 0_n$, let $\mathcal{D}_s$ be the cut vector specified

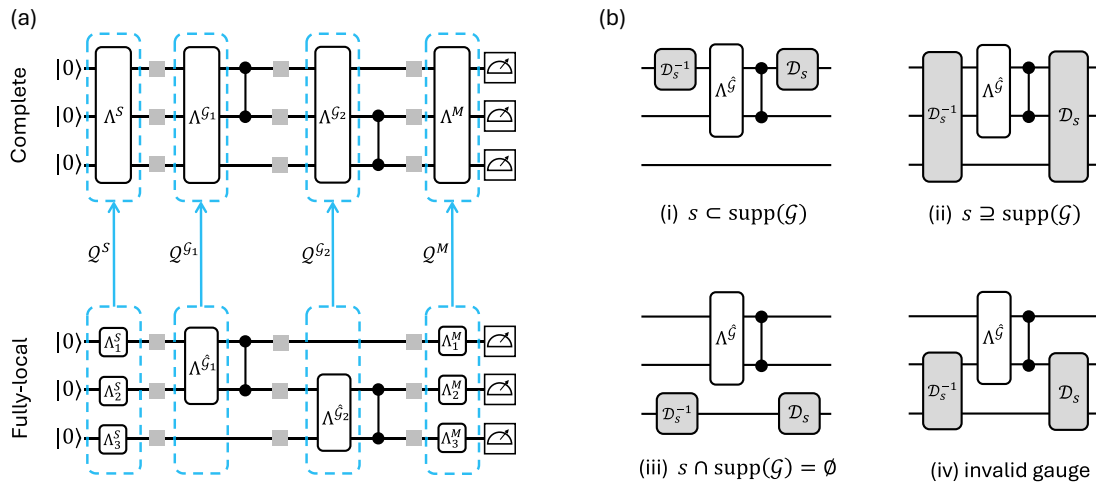


FIG. 3. (a) Illustration of a fully local noise model (bottom) and how it is embedded into a complete noise model (above). Note that the state prep., meas., and each layer of gates have an independent embedding map, which means this is a layer-uncorrelated noise model. (b) Illustration of the subsystem depolarizing gauges (denoted by $\mathcal{D}_s$) acting on a gate with fully local noise. Here, (i)–(iii) shows the three types of valid SDG's as given by Theorem 10(b), out of which only case (i) can non-trivially changes the noise channel. (iv) shows a SDG that is not an allowed gauge transformation (when the gate has a fully-connected pattern transfer subgraph). Intuitively, it propagates the two-qubit noise channel into a three-qubit channel, violating the fully-local assumptions.

by the vertex partition

$$V_0 = \{u : u_s = 0_{|s|}\} \cup \{v_R\}, \quad V_0^c = \{u : u_s \neq 0_{|s|}\}.$$

We call such $\mathfrak{d}_s$ a **subsystem depolarizing gauge** (SDG).

To see why $\mathfrak{d}_s$ is called a subsystem depolarizing gauge, let us first write down its expansion on the standard basis:

$$\begin{aligned} e_u^S \cdot \mathfrak{d}_s &= \mathbb{1}[u_s \neq 0_{|s|}], \\ e_u^M \cdot \mathfrak{d}_s &= -\mathbb{1}[u_s \neq 0_{|s|}], \\ e_a^G \cdot \mathfrak{d}_s &= \mathbb{1}[\text{pt}(\mathcal{G}(a))_s \neq 0_{|s|}] - \mathbb{1}[\text{pt}(a)_s \neq 0_{|s|}], \\ \forall u \in \{0, 1\}^n \setminus 0_n, \quad \forall \mathcal{G} \in \mathfrak{G}, \quad \forall a \in \mathbf{P}^n \setminus I_n, \end{aligned} \quad (38)$$

which can be obtained using the definition of cut vectors and the PTG. Now, consider a transformation of noise parameters given by $\mathbf{x} \mapsto \mathbf{x} + \eta \mathfrak{d}_s$. It is not hard to verify this corresponds to the following gauge transformation:

$$\begin{aligned} \tilde{\rho}_0 &\mapsto \mathcal{D}_{s,\eta}(\tilde{\rho}_0), \quad \tilde{E}_j \mapsto \mathcal{D}_{s,\eta}^{-1}(\tilde{E}_j), \\ \tilde{\mathcal{G}}_s &\mapsto \mathcal{D}_{s,\eta} \circ \tilde{\mathcal{G}} \circ \mathcal{D}_{s,\eta}^{-1}, \end{aligned} \quad (39)$$

where $\mathcal{D}_{s,\eta} := \sum_a e^{-\eta \mathbb{1}[a_s \neq I_{|s|}]} |\sigma_a\rangle\langle\sigma_a|$ is a partially depolarizing channel (with strength parameter η) acting on the subsystem specified by the index set s . For notational convenience, we define

$$\mathcal{D}_s := \mathcal{D}_{s,1} = \sum_a e^{-\mathbb{1}[a_s \neq I_{|s|}]} |\sigma_a\rangle\langle\sigma_a|, \quad (40)$$

and call $\mathcal{D}_s$ simply as the subsystem depolarizing channel on s .

Lemma 8. The set of SDGs $\{\mathfrak{d}_s\}_{s \neq 0_n}$ forms a basis for the gauge space T .

Proof. Since $\#\{\mathfrak{d}_s\}_s = \dim(T) = 2^n - 1$ (recall that the PTG is strongly connected thanks to the root vertex), we only need to show the completeness of $\{\mathfrak{d}_s\}_s$. It is sufficient to show that $\{\mathfrak{d}_s\}_s$ can express the canonical cut basis η_z generated by cutting off a single vertex $V_0 = \{v_z\}$ for all non-zero bit string z . (Note that the root node gives a redundant degree of freedom and can thus be ignored.) By definition, we have

$$\mathfrak{d}_s = - \sum_z \mathbb{1}[z_s \neq 0] \eta_z. \quad (41)$$

For two n -bit string s, t we write $t \leq s$ if $t_i = 0$ for any index i such that $s_i = 0$. Let $\bar{s}$ denote the bitwise flip of

s (e.g., $\overline{1101} = 0010$). We claim the following relation holds:

$$\eta_z = \sum_{s \neq \mathbf{0}: \bar{z} \leq s} (-1)^{|s| - |\bar{z}|} \mathfrak{d}_s, \quad \forall z \neq \mathbf{0}. \quad (42)$$

This is basically a consequence of the inclusion-exclusion principle. To see this, thanks to Eq. (41),

$$\begin{aligned} \text{RHS} &= - \sum_{s \neq \mathbf{0}: \bar{z} \leq s} (-1)^{|s| - |\bar{z}|} \sum_{t: t_s \neq 0} \eta_t \\ &= - \sum_t \sum_{s: \bar{z} \leq s, s \not\leq \bar{t}} (-1)^{|s| - |\bar{z}|} \eta_t \\ &= \sum_t \sum_{s: \bar{z} \leq s \leq \bar{t}} (-1)^{|s| - |\bar{z}|} \eta_t \\ &= \sum_t \mathbb{1}[\bar{z} \leq \bar{t}] \sum_{k=0}^{|\bar{t}| - |\bar{z}|} (-1)^k \binom{|\bar{t}| - |\bar{z}|}{k} \eta_t \\ &= \sum_t \mathbb{1}[\bar{z} \leq \bar{t}] \mathbb{1}[|\bar{t}| - |\bar{z}| = 0] \eta_t \\ &= \eta_z. \end{aligned} \quad (43)$$

The order change of sum in the second line uses the fact that $t_s \neq 0$ is equivalent to $s \not\leq \bar{t}$, and that $\mathbf{0} \leq \bar{t}$ for any t , so we can drop the constraint $s \neq \mathbf{0}$. The third line uses the following:

$$\begin{aligned} \sum_{s: \bar{z} \leq s \leq \bar{t}} (-1)^{|s| - |\bar{z}|} + \sum_{s: \bar{z} \leq s, s \not\leq \bar{t}} (-1)^{|s| - |\bar{z}|} \\ = \sum_{s: \bar{z} \leq s} (-1)^{|s| - |\bar{z}|} = \sum_{k=0}^{n - |\bar{z}|} \binom{n - |\bar{z}|}{k} (-1)^k = 0. \end{aligned} \quad (44)$$

Note that $z \neq \mathbf{0}$ implies $n - |\bar{z}| > 0$. The fifth line uses the Binomial theorem. This proves Eq. (42), and thus establishes the fact that the SDG's $\{\mathfrak{d}_s\}_s$ form a basis for T and U . ■

In fact, the gauge space of the fully local noise model is spanned by a subset of the SDG's. We have the following results:

Theorem 10. Given a layer-uncorrelated noise model (X_R, Q) .

- (a) If the model satisfies fully local state preparation noise, then

$$\{\mathbf{z} \in T : \mathbf{z}^S \in \text{Im } Q^S\} = \text{span}\{\mathfrak{d}_s : |s| = 1\}. \quad (45)$$

Similarly for the measurement noise.

- (b) If the model satisfies fully local noise for some gate $\mathcal{G} \in \mathfrak{G}$, then

$$\{\mathbf{z} \in T : \mathbf{z}^{\mathcal{G}} \in \text{Im } \mathcal{Q}^{\mathcal{G}}\} \supseteq \text{span}\{\mathfrak{d}_s : s \subset \text{supp}(\mathcal{G}) \text{ or } s \supseteq \text{supp}(\mathcal{G}) \text{ or } s \cap \text{supp}(\mathcal{G}) = \emptyset\}. \quad (46)$$

Furthermore, the two sides are equal if $\hat{\mathcal{G}}$ induces a connected pattern transfer subgraph.

- (c) If the model is fully local, its embedded gauge space is given by

$$\mathcal{Q}(T_R) = \text{span}\{\mathfrak{d}_s : |s| = 1\}. \quad (47)$$

And thus $\dim T_R = \dim \mathcal{Q}(T_R) = n$.

Here, the pattern transfer subgraph of $\hat{\mathcal{G}}$ is a $|\mathcal{G}|$ -qubit PTG with the only gate being $\hat{\mathcal{G}}$, and with the SPAM (root) node removed. Examples of two-qubit gates that have connected pattern transfer subgraph include CZ, *i*SWAP, but not SWAP [18], see Fig. 11. As another example, the pattern transfer subgraph for $\text{CZ} \otimes \text{CZ}$ is not connected. The allowed and forbidden gauge transformation for the fully local gate noise are depicted in Fig. 3(b).

Theorem 10(c) characterizes the learnability of fully local Pauli noise models. In words, the embedded gauge space $\mathcal{Q}(T_R)$ is spanned exactly by those n single-qubit depolarizing gauges. One can also write down the reduced gauge space as $T_R = \text{span}\{\mathcal{Q}^{-1}(\mathfrak{d}_s) : |s| = 1\}$, where each $\mathcal{Q}^{-1}(\mathfrak{d}_s)$ is well-defined [74] and can be efficiently calculated using the inverse transformation given in Lemma 7. The learnable space L_R can be obtained by, e.g., computing the orthogonal complement of T_R within X_R using standard linear algebra procedures. We will come back to the construction of L_R later.

Now we present the proof for Theorem 10(a), the direct part of Theorem 10(b), and Theorem 10(c). The converse part of Theorem 10(b) is slightly more involved and is postponed to Appendix C for clarity.

Proof of Theorem 10(a). Let $\pi^S : X \mapsto X^S$ be the projection map $\pi^S(\mathbf{x}) = \mathbf{x}^S$, i.e., restriction to the state preparation parameters. We first show $\text{LHS} \supseteq \text{RHS}$, for which we only need to prove that $\pi^S(\mathfrak{d}_{1_j}) \in \text{Im } \mathcal{Q}^S$ for all $j \in [n]$. Indeed, we have that,

$$\pi^S(\mathfrak{d}_{1_j}) = \sum_u \mathbb{1}[u_j \neq 0] \mathbf{e}_u^S = \mathcal{Q}^S(\mathfrak{d}_j^S). \quad (48)$$

The two equations are by Definitions 19 and 18, respectively. Thus we have $\text{LHS} \supseteq \text{RHS}$.

Conversely, notice that π^S is an invertible map when restricted to T . To see this, consider the canonical cut basis $\{\eta_u : u \neq 0_n\}$ for T such that each η_u is induced by the vertex partition $\mathbf{V} = \mathbf{V}_0 \cup \mathbf{V}_0^c$ where $\mathbf{V}_0 = \{u\}$. We have

$\pi^S(\eta_u) = -\mathbf{e}_u^S$. This means π^S is a basis transform between T and X^S , thus invertible. Therefore,

$$\begin{aligned} \dim(\{\mathbf{z} \in T : \pi^S(\mathbf{z}) \in \text{Im } \mathcal{Q}^S\}) \\ = \dim(\text{Im } \mathcal{Q}^S) \leq \dim(X_R^S) = n, \end{aligned} \quad (49)$$

where the first equation uses the invertibility of π^S restricted to T . This means the dimension of LHS is no larger than RHS. Consequently, they must be equal. The proof works similarly on the measurement side. ■

Proof of Theorem 10(b). Let $\pi^{\mathcal{G}} : X \mapsto X^{\mathcal{G}}$ be the projection map $\pi^{\mathcal{G}}(\mathbf{x}) = \mathbf{x}^{\mathcal{G}}$. Recall from Eq. (38) that for all $a \neq I_n$,

$$\mathbf{e}_a^{\mathcal{G}} \cdot \mathfrak{d}_s = \mathbb{1}[\text{pt}(\mathcal{G}(a))_s \neq 0_{|s|}] - \mathbb{1}[\text{pt}(a)_s \neq 0_{|s|}].$$

- (1) If $s \cap \text{supp}(\mathcal{G}) = \emptyset$: we have $\text{pt}(\mathcal{G}(a))_s = \text{pt}(a)_s$ and thus $\mathbf{e}_a^{\mathcal{G}} \cdot \mathfrak{d}_s = 0$ for all a . Therefore, we trivially have $\pi^{\mathcal{G}}(\mathfrak{d}_s) \in \text{Im } \mathcal{Q}^{\mathcal{G}}$.
- (2) If $s \supseteq \text{supp}(\mathcal{G})$: we have $\text{pt}(\mathcal{G}(a))_s = 0_{|s|} \Leftrightarrow \text{pt}(a)_s = 0_{|s|}$ because $\hat{\mathcal{G}}(a_{\text{supp}(\mathcal{G})}) = I_{|\mathcal{G}|}$ if and only if $a_{\text{supp}(\mathcal{G})} = I_{|\mathcal{G}|}$. Thus $\mathbf{e}_a^{\mathcal{G}} \cdot \mathfrak{d}_s = 0$ and $\pi^{\mathcal{G}}(\mathfrak{d}_s) \in \text{Im } \mathcal{Q}^{\mathcal{G}}$ trivially holds.
- (3) If $s \subset \text{supp}(\mathcal{G})$: choose an $\mathbf{r}^{\mathcal{G}} \in X_R^{\mathcal{G}}$ such that

$$\mathfrak{d}_b^{\mathcal{G}} \cdot \mathbf{r}^{\mathcal{G}} = \mathbb{1}[\text{pt}(\hat{\mathcal{G}}(b))_s \neq 0_{|s|}] - \mathbb{1}[\text{pt}(b)_s \neq 0_{|s|}]. \quad (50)$$

Then we can see that

$$\begin{aligned} \mathbf{e}_a^{\mathcal{G}} \cdot \mathcal{Q}^{\mathcal{G}}(\mathbf{r}^{\mathcal{G}}) &= \mathbb{1}[\text{pt}(\hat{\mathcal{G}}(a_{\text{supp}(\mathcal{G})}))_s \neq 0_{|s|}] \\ &\quad - \mathbb{1}[\text{pt}(a_{\text{supp}(\mathcal{G})})_s \neq 0_{|s|}] = \mathbf{e}_a^{\mathcal{G}} \cdot \mathfrak{d}_s. \end{aligned} \quad (51)$$

This means $\pi^{\mathcal{G}}(\mathfrak{d}_s) = \mathcal{Q}^{\mathcal{G}}(\mathbf{r}^{\mathcal{G}})$. Thus we have $\pi^{\mathcal{G}}(\mathfrak{d}_s) \in \text{Im } \mathcal{Q}^{\mathcal{G}}$.

Combining all the three cases and using linearity, we conclude that

$$\{\mathbf{z} \in T : \mathbf{z}^{\mathcal{G}} \in \text{Im } \mathcal{Q}^{\mathcal{G}}\} \supseteq \text{span}\{\mathfrak{d}_s : s \subset \text{supp}(\mathcal{G}) \text{ or } s \supseteq \text{supp}(\mathcal{G}) \text{ or } s \cap \text{supp}(\mathcal{G}) = \emptyset\}. \quad (52)$$

Conversely, the two sides are equal if $\hat{\mathcal{G}}$ induces a connected pattern transfer subgraph, which is proven in Appendix C. The proof works by establishing an upper bound on the dimension of the LHS, and then show by counting that this bound equals the dimension of the RHS. ■

We make a few remarks regarding the proof of Theorem 10(b). First, it is clear from the proof that $\text{span}\{\mathfrak{d}_s : s \subset \text{supp}(\mathcal{G})\}$ are gauges that non-trivially change the noise parameters of $\mathcal{G}$, while $\text{span}\{\mathfrak{d}_s : s \supseteq \text{supp}(\mathcal{G}) \text{ or } s \cap \text{supp}(\mathcal{G}) = \emptyset\}$ are gauges that “pass through” $\mathcal{G}$ leaving gate noise parameters unchanged; Second, the proof for Eq. (52) taking equality crucially relies on the connectivity of the pattern transfer subgraph of $\widehat{\mathcal{G}}$. If this is not satisfied, there could be more gauges that can trivially pass through $\mathcal{G}$, which needs to be analyzed *ad hoc* using Theorem 8.

Proof of Theorem 10(c). Thanks to Lemma 6, $\mathcal{Q}(T_R)$ is given by the intersection of Eq. (45) for SPAM and Eq. (46) for all $\mathcal{G} \in \mathfrak{G}$. It is easy to see that Eq. (45) is always a subset of Eq. (46) for any $\mathcal{G} \in \mathfrak{G}$, thus their intersection is simply given by Eq. (45), the single-qubit SDGs. ■

C. Learning algorithm for fully local model

Having characterized the reduced gauge space of the fully local noise model, we turn to the task of designing efficient learning experiments. Following the discussion in

Sec. IV, the key is to find an appropriate reduced cycle basis for L_R . Let us start with the first type of construction which gives the simplest experimental design (but does not care about relative precision estimation). Recall the rooted cycle basis defined in Eq. (18),

$$\mathbf{B} := \{\mathbf{e}_u^S + \mathbf{e}_u^M : u \neq 0_n\} \cup \{\mathbf{e}_{\text{pt}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{pt}(\mathcal{G}(a))}^M : \mathcal{G} \in \mathfrak{G}, a \neq I_n\}. \quad (53)$$

Consider the following subset of $\mathbf{B}$, denoted as $\mathbf{B}'$,

$$\mathbf{B}' := \{\mathbf{e}_{1_j}^S + \mathbf{e}_{1_j}^M : j \in [n]\} \cup \{\mathbf{e}_{\text{pt}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{pt}(\mathcal{G}(a))}^M : \mathcal{G} \in \mathfrak{G}, \text{supp}(a) \subseteq \text{supp}(\mathcal{G}), a \neq I\}. \quad (54)$$

Proposition 1. Let $\mathbf{B}_R := \mathcal{Q}^T(\mathbf{B}')$. Then, $\mathbf{B}_R$ forms a reduced cycle basis for L_R .

Proof. We have

$$\begin{aligned} \mathbf{B}_R &:= \mathcal{Q}^T \left(\{\mathbf{e}_u^S + \mathbf{e}_u^M : |u| = 1\} \cup \{\mathbf{e}_{\text{pt}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{pt}(\mathcal{G}(a))}^M : \mathcal{G} \in \mathfrak{G}, \text{supp}(a) \subseteq \text{supp}(\mathcal{G}), a \neq I_n\} \right) \\ &= \{\boldsymbol{\vartheta}_j^S + \boldsymbol{\vartheta}_j^M : j \in [n]\} \cup \left\{ \sum_{j \in \text{supp}(\mathcal{G})} (\mathbb{1}[\text{pt}(b)_j \neq 0] \boldsymbol{\vartheta}_j^S + \mathbb{1}[\text{pt}(\widehat{\mathcal{G}}(b))_j \neq 0] \boldsymbol{\vartheta}_j^M) + \boldsymbol{\vartheta}_b^G : \mathcal{G} \in \mathfrak{G}, b \neq I_{|\mathcal{G}|} \right\}. \end{aligned} \quad (55)$$

Recall that $\widehat{\mathcal{G}}$ is $\mathcal{G}$ restricted to its support. It is not hard to see that $\mathbf{B}_R$ forms a linearly-independent set. Meanwhile, $|\mathbf{B}_R| = n + \sum_{\mathcal{G} \in \mathfrak{G}} (4^{|\mathcal{G}|} - 1) = \dim(X_R) - \dim(T_R) = \dim(L_R)$. This means $\mathbf{B}_R$ is indeed a basis for L_R . ■

Following Algorithm 1, the following set of experiments is sufficient to learn the noise model:

- (1) For $j \in [n]$, prepare $\tilde{\rho}_0$ and measure $\tilde{Z}$ on the j th qubit.
- (2) For $\mathcal{G} \in \mathfrak{G}$, $b \in \mathbf{P}^{|\mathcal{G}|} \setminus I_{|\mathcal{G}|}$, prepare $\tilde{\rho}_b$, apply $\tilde{\mathcal{G}}$, and measure $\tilde{P}_{\widehat{\mathcal{G}}(b)}$ on the support of $\mathcal{G}$.

The total number of experiments is $n + \sum_{\mathcal{G} \in \mathfrak{G}} (4^{|\mathcal{G}|} - 1)$. If each $\mathcal{G}$ acts non-trivially only on a constant number of qubits, the number of experiments needed is $O(n)$, and is efficient to conduct. We also remark that, here we have not taken advantages of the fact that multiple observables can be extracted from the same experiments. In practice, the sample complexity can thus be further reduced.

To learn the gate noise parameters to relative precision, one needs to construct a reduced cycle basis for the gate noise parameter space $L_R^G = \mathcal{Q}^T(L \cap X^G)$. We find this problem challenging in general. One reason is that some reduced cycle can only be represented using parameters of many different gates. However, if every $\mathcal{G} \in \mathfrak{G}$ is supported non-trivially on at most two qubits, we do have an efficient algorithms to construct such a basis. The key idea is that there are only three types of reduced cycles that involves more than one gate in this case. Details are presented in Appendix D.

VI. APPLICATIONS TO QUASI-LOCAL NOISE MODEL

We have presented a comprehensive treatment on learning the fully local Pauli noise model. In practice, the effect of spatially correlated noise can sometimes be non-negligible, both for SPAM [75] and for gates [76,77]. It is thus desirable to study a quasi-local noise model that takes a finite spatial correlation (e.g., nearest-neighbor) into consideration while keeping the total number of parameters

manageable. Such models have been studied in the literature of Pauli noise learning [10,11,15–17]. In this section, we will introduce the quasi-local model in the framework of gate set Pauli noise learning, provide results on its learnability, and discuss efficient learning algorithms. As a specific example, we will apply our theory to a noise model that has been intensively studied in recent experiments [17,49] and obtain a self-consistent algorithm for learning the noise model, which has not been previously achieved.

A. Basic definitions

We first introduce some notations that are needed for defining the quasi-local noise model. For $a, b \in \mathbf{P}^n$, we write $b \triangleleft a$ if $b_i \neq I \Rightarrow a_i = b_i$ for all $i \in [n]$, calling b majorized by a . For any $S \subseteq [n]$, we write $a \sim S$ if $a_i \neq I \Rightarrow i \in S$. For any $\mathbb{S} \subseteq 2^{[n]}$, we write $a \sim \mathbb{S}$ if $a \sim S$ for at least one $S \in \mathbb{S}$.

Let Ω be a subset of $2^{[n]}$ that satisfies $\forall v \in \Omega, \forall \emptyset \subsetneq \mu \subseteq v, \mu \in \Omega$. We call such Ω a *factor set*. We also use Ω_* to denote the set of maximal factors within Ω , i.e., those that do not belong to any other factor. As an example, $\Omega = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}\}$ is a valid factor set, with the corresponding $\Omega_* = \{\{1,2\}, \{2,3\}\}$. Obviously, Ω, Ω_* uniquely determines each other.

Definition 20. An n -qubit Pauli channel Λ is Ω -local if its Pauli eigenvalues satisfies

$$\lambda_a = \prod_{b \triangleleft a, b \sim \Omega} \exp(-r_b), \quad \forall a \neq I_n, \quad (56)$$

for some $\{r_b \in \mathbb{R} : b \sim \Omega, b \neq I_n\}$ and a factor set Ω .

For notational simplicity, we sometimes omit the requirement that $b \neq I_n$ and set $r_{I_n} = 0$.

A few remarks regarding this definition: First, all n -qubit Pauli channels (with all Pauli eigenvalues being positive) are $2^{[n]} \setminus \{\emptyset\}$ -local. Note that we do not require r_b to be non-negative. Second, this definition basically says that the eigenvalues $\{\lambda_a\}_a$ factorize according to a Markov random field (MRF) [78] specified by the factor decomposition Ω . This is opposed to Ref. [10] where Pauli channels with error rates $\{p_a\}_a$ factorized according to an MRF are considered. It is unclear whether these two models are equivalent, even approximately. Alternatively, our definition of Ω -local is equivalent to requiring that Λ can be expressed as a concatenation of Pauli diagonal maps [79] acting on each subsystem from Ω , as proven in Ref. [16] (In fact, this is the picture we will use in most figures shown in this section.) One can also show that the sparse Pauli-Lindblad model used in Refs. [17,49,80] is equivalent to our definition. Finally, throughout this paper we assume Ω is known *a priori*, from, e.g., knowledge about the device topology, leaving the topic of structure

learning (see, e.g., Ref. [64]) for future study. We provide more details about the quasi-local Pauli noise models in Appendix E.

B. Learnability of the quasi-local model

Note that, in terms of x_a , Definition 20 gives that $x_a = \sum_{b \triangleleft a, b \sim \Omega} r_b$. This motivates the following definition of a quasi-local Pauli noise model. Note that, we have taken into account that the SPAM noise channels are symmetric Pauli channels (i.e., eigenvalues only depend on the Pauli pattern). An example of the quasi-local noise model is given in Fig. 4.

Definition 21 (Quasi-local noise). Consider a layer-uncorrelated noise model (X_R, Q) :

- (1) If the state preparation noise parameters are given by $\{r_v^S : v \in \Omega^S\}$, where Ω^S is a factor set, such that

$$\mathcal{Q}(\vartheta_v^S) = \sum_{\mu \neq \emptyset} \mathbb{1}[v \subseteq \mu] e_\mu^S, \quad \forall v \in \Omega^S, v \neq \emptyset, \quad (57)$$

we say the model has Ω^S -local state preparation noise. Similar for the measurement.

- (2) If the noise parameters for some $\mathcal{G} \in \mathfrak{G}$ are given by $\{r_b^{\mathcal{G}} : b \sim \Omega^{\mathcal{G}}, b \neq I_n\}$, where $\Omega^{\mathcal{G}}$ is a factor set, such that

$$\mathcal{Q}(\vartheta_b^{\mathcal{G}}) = \sum_{a \neq I_n} \mathbb{1}[b \triangleleft a] e_a^{\mathcal{G}}, \quad \forall b \sim \Omega^{\mathcal{G}}, b \neq I_n. \quad (58)$$

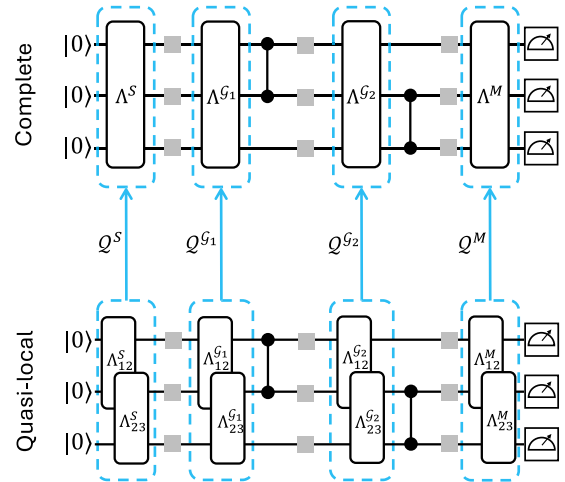


FIG. 4. Example of a three-qubit quasi-local noise model and the embedding into a complete model. Here, the noise channels for state prep., meas., and all gates are assumed to be Ω -local where $\Omega_* = \{\{1,2\}, \{2,3\}\}$. In other words, $\Lambda^S = \Lambda_{12}^S \circ \Lambda_{23}^S$ where $\Lambda_{12}^S, \Lambda_{23}^S$ are Pauli diagonal maps supporting on $\{1,2\}, \{2,3\}$, respectively. Similar for other Pauli noise channels.

We say the model has **$\Omega^{\mathcal{G}}$ -local noise for $\mathcal{G}$** .

- (3) If the state preparation, measurements, and all layers have quasi-local noise with locality parameters $\{\Omega^S, \Omega^M, \Omega^{\mathcal{G}} : \mathcal{G} \in \mathfrak{G}\}$, we say the model is quasi-local with the corresponding parameters.

The following lemma describes the transformation between the reduced and the complete parameters within a quasi-local model.

Lemma 9. If $(X_R, \mathcal{Q})$ satisfies Ω^S -local state preparation noise, then

$$\begin{cases} x_{\mu}^S = \sum_{\nu \subseteq \mu, \nu \in \Omega^S} r_{\nu}^S, & \forall \mu \in 2^{[n]} \setminus \emptyset. \\ r_{\nu}^S = \sum_{\mu \subseteq \nu, \mu \in \Omega^S} (-1)^{|\nu| - |\mu|} x_{\mu}^S, & \forall \nu \in \Omega^S. \end{cases} \quad (59)$$

Similar for the measurement side; If $(X_R, \mathcal{Q})$ satisfies $\Omega^{\mathcal{G}}$ -local noise for a gate $\mathcal{G}$, then

$$\begin{cases} x_a^{\mathcal{G}} = \sum_{b \triangleleft a, b \sim \Omega^{\mathcal{G}}} r_b^{\mathcal{G}}, & \forall a \in \mathbb{P}^n \setminus I_n. \\ r_b^{\mathcal{G}} = \sum_{a \triangleleft b, a \sim \Omega^{\mathcal{G}}} (-1)^{w(b) - w(a)} x_a^{\mathcal{G}}, & \forall b \sim \Omega^{\mathcal{G}}. \end{cases} \quad (60)$$

Proof. The transformation from r 's to x 's can be obtained by expanding $\mathbf{x} = \mathcal{Q}(\mathbf{r})$ in the standard basis and compare coefficients. The inverse transformation can be verified by substitution. As a side comment, these pairs of transformation are known as the Möbius transformation. They are also used in Ref. [16] to study Pauli noise learning in the logical level. ■

As a sanity check, one can retrieve the fully local model in Definition 18 by setting

$$\begin{aligned} \Omega^S &= \Omega^M = \{\{j\} : j \in [n]\}, \\ \Omega^{\mathcal{G}} &= \{\nu : \emptyset \neq \nu \subseteq \text{supp}(\mathcal{G})\}, \quad \forall \mathcal{G} \in \mathfrak{G}. \end{aligned} \quad (61)$$

Now, we make two important remarks about the gate noise models. First, we will take a more general approach to deal with parallel gates than the fully local case. Instead of letting each $\mathcal{G}$ to have a constantly small support, we allow it to be a tensor product of many few-qubit gates, i.e., $\mathcal{G} = \prod_{i=1}^M \hat{\mathcal{G}}_i$ where $\{\hat{\mathcal{G}}_i\}$ are Clifford gates supported non-trivially on disjoint sets of qubits. Second, note that a Clifford gate affected by Pauli noise can be modeled in two equivalent ways: whether the noise channel happens before or after the gate, i.e.,

$$\tilde{\mathcal{G}} = \mathcal{G} \circ \Lambda = \Lambda' \circ \mathcal{G}, \quad \text{where } \Lambda' = \mathcal{G} \circ \Lambda \circ \mathcal{G}^{\dagger}. \quad (62)$$

Throughout this work, we have been adopting the noise-before-gate convention, but now it is worth looking at

the other convention. Specifically, when Λ is an Ω -local Pauli channel, Λ' might not be Ω -local in general. This raised the following question: is there a physical reason to believe the gate noise has certain locality constraint in the noise-before-gate convention, but not in the noise-after-gate convention? To the best of our knowledge, no physical justification of this kind has been established in the literature. To circumvent the dilemma, we propose to study a specific type of locality assumption:

Definition 22. An Ω -local noise model is called **$\mathcal{G}$ -covariant**, if for any Ω -local Pauli channel Λ , it holds that $\mathcal{G} \circ \Lambda \circ \mathcal{G}^{\dagger}$ and $\mathcal{G}^{\dagger} \circ \Lambda \circ \mathcal{G}$ are also Ω -local Pauli channels.

Since a $\mathcal{G}$ -covariant noise model has the same locality constraints in either convention, it is much easier to justify their physical relevance. An example of covariant vs invariant noise model is given in Fig. 5. When $\mathcal{G}$ represents a parallel application of many Clifford gates with disjoint supports, we have the following sufficient condition for a noise model to be $\mathcal{G}$ -covariant:

Lemma 10. Given $\mathcal{G} = \prod_{i=1}^M \hat{\mathcal{G}}_i$ where $\{\hat{\mathcal{G}}_i\}$ are Clifford gates supported non-trivially on disjoint sets of qubits. Define the extended support map $\Xi_{\mathcal{G}} : [n] \mapsto [n]$ as

$$\Xi_{\mathcal{G}}(s) := \left(\bigcup_{\text{supp}(\hat{\mathcal{G}}_i) \cap s \neq \emptyset} \text{supp}(\hat{\mathcal{G}}_i) \right) \cup s. \quad (63)$$

For an Ω -local noise model, if $\forall s \in \Omega$, $\Xi_{\mathcal{G}}(s) \in \Omega$, then Ω is $\mathcal{G}$ -covariant.

Clearly, for a Pauli P with support s , the support of $\mathcal{G}(P)$ cannot exceed $\Xi_{\mathcal{G}}(s)$. Lemma 10 roughly says that if for all Pauli $P \sim \Omega$, one has $\mathcal{G}(P) \sim \Omega$, then Ω must be $\mathcal{G}$ -covariant. The proof is given in Appendix E.

We are ready to present our main results on the learnability of quasi-local Pauli noise model. Perhaps not surprisingly, the gauge degrees of freedom can still be described using the SDG's in Definition 19, as follows:

Theorem 11. Given a layer-uncorrelated noise model $(X_R, \mathcal{Q})$

- (a) If the model has Ω^S -local state preparation noise, then

$$\{z \in T : z^S \in \text{Im } \mathcal{Q}^S\} = \text{span}\{\mathfrak{d}_{\nu} : \nu \in \Omega^S\}. \quad (64)$$

Similarly for the measurement noise.

- (b) If the model has $\Omega^{\mathcal{G}}$ -local noise for some $\mathcal{G} \in \mathfrak{G}$ and is $\mathcal{G}$ -covariant, then

$$\begin{aligned} \{z \in T : z^{\mathcal{G}} \in \text{Im } \mathcal{Q}^{\mathcal{G}}\} \\ \supseteq \text{span}\{\mathfrak{d}_{\nu} : \nu \in \Omega^{\mathcal{G}} \text{ or } \Xi_{\mathcal{G}}(\nu) = \nu\}. \end{aligned} \quad (65)$$

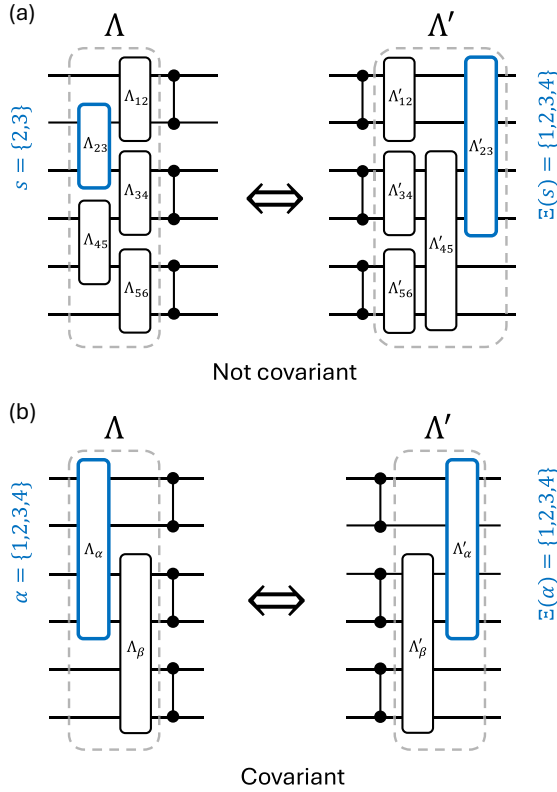


FIG. 5. Examples of quasi-local gate noise models that are incovariant and covariant, respectively. Here we consider a six-qubit system and let $\mathcal{G} = \text{CZ}^{\otimes 3}$. (a) The nearest-neighbor two-local noise model ($\Omega^* = \{\{1, 2\}, \{2, 3\}, \dots, \{5, 6\}\}$) which is not $\mathcal{G}$ -covariant. Indeed, if we commute the noise channel from before- $\mathcal{G}$ (left) to after- $\mathcal{G}$ (right), certain two-body noise terms (e.g., Λ_{23}) will generically be propagated to be four-body, making the whole noise channel no longer Ω -local; (b) A four-local noise model ($\Omega^* = \{\alpha = \{1, 2, 3, 4\}, \beta = \{3, 4, 5, 6\}\}$) which is $\mathcal{G}$ -covariant. Indeed, commuting the noise channel from before- $\mathcal{G}$ to after- $\mathcal{G}$ preserves the Ω -local structure. Note that all Pauli diagonal maps commute with each other. The bold blue boxes demonstrate how $\Xi_{\mathcal{G}}$ acts on different set of indexes.

- (c) If $\Omega^S \subseteq \Omega^M$, $\Omega^{\mathcal{G}}$ is $\mathcal{G}$ covariant for all $\mathcal{G} \in \mathfrak{G}$, and it holds that $\forall v \in \Omega^S, \forall \mathcal{G} \in \mathfrak{G}$, either $v \in \Omega^{\mathcal{G}}$ or $\Xi_{\mathcal{G}}(v) = v$, then the embedded gauge space for the whole model is given by

$$\mathcal{Q}(T_R) = \text{span}\{\mathfrak{d}_v : v \in \Omega^S\}. \quad (66)$$

Theorem 11 (c) confirms that, given certain “goodness” assumptions of the quasi-local noise model, the embedded gauge space is indeed spanned by a subset of SDG’s consistent with the factor sets. Similar to the fully local case, one can efficiently compute the reduced gauge space $T_R = \text{span}\{\mathcal{Q}^{-1}(\mathfrak{d}_v) : v \in \Omega^S\}$ by applying the Möbius inverse transformation given in Lemma 9.

Proof of Theorem 11 (a). Let $\pi^S : X \mapsto X^S$ be the projection map $\pi^S(x) = x^S$. Recall from the proof of

Theorem 10 (a) that π^S is an isomorphism between T and X^S . We claim the following two linear spaces are equal.

$$\text{span}\{\mathcal{Q}^S(\mathfrak{d}_v^S) : v \in \Omega^S\} = \text{span}\{\pi^S(\mathfrak{d}_v) : v \in \Omega^S\}. \quad (67)$$

Now we prove this claim. By definition, for any $v \in \Omega^S$ we have

$$\mathcal{Q}^S(\mathfrak{d}_v^S) = \sum_{\mu \neq \emptyset} \mathbb{1}[v \subseteq \mu] e_\mu^S =: \alpha_v, \quad (68)$$

$$\pi^S(\mathfrak{d}_v) = \sum_{\mu \neq \emptyset} \mathbb{1}[v \cap \mu \neq \emptyset] e_\mu^S =: \beta_v. \quad (69)$$

Since both $\{\alpha_v\}, \{\beta_v\}$ have the same cardinality, and we know $\{\beta_v\}$ are linearly-independent thanks to the linear independence of the SDGs and π^S being an isomorphism from T to X^S , we just need to show $\{\alpha_v\}$ can linearly represent $\{\beta_v\}$. We claim the following holds:

$$\beta_v = - \sum_{v' \in \Omega^S} \mathbb{1}[v' \subseteq v] (-1)^{|v'|} \alpha_{v'}, \quad \forall v \in \Omega^S. \quad (70)$$

Indeed, expanding the RHS yields,

$$\begin{aligned} \text{RHS} &= - \sum_{\mu \neq \emptyset} e_\mu^S \sum_{v' \in \Omega^S} \mathbb{1}[v' \subseteq v] \mathbb{1}[v' \subseteq \mu] (-1)^{|v'|} \\ &= - \sum_{\mu \neq \emptyset} e_\mu^S \sum_{v' \in \Omega^S} \mathbb{1}[v' \subseteq v \cap \mu] (-1)^{|v'|} \\ &= - \sum_{\mu \neq \emptyset} e_\mu^S \sum_{\substack{v' \subseteq v \cap \mu, \\ v' \neq \emptyset}} (-1)^{|v'|} \\ &= \sum_{\mu \neq \emptyset} e_\mu^S \mathbb{1}[v \cap \mu \neq \emptyset] = \beta_v, \end{aligned} \quad (71)$$

where the third line uses the defining property of Ω^S that for any $v \in \Omega^S$, any non-empty $v' \subseteq v$ also satisfies $v' \in \Omega^S$; The last line uses the binomial theorem. This completes the proof of Eq. (67). Consequently, we have

$$\begin{aligned} \{\mathbf{z} \in T : \mathbf{z}^S \in \text{Im} \mathcal{Q}^S\} &= \{\mathbf{z} \in T : \pi^S(\mathbf{z}) \in \text{span}\{\mathcal{Q}^S(\mathfrak{d}_v^S) : v \in \Omega^S\}\} \\ &= \{\mathbf{z} \in T : \pi^S(\mathbf{z}) \in \text{span}\{\pi^S(\mathfrak{d}_v) : v \in \Omega^S\}\} \\ &= \text{span}\{\mathfrak{d}_v : v \in \Omega^S\}. \end{aligned} \quad (72)$$

The second line uses Eq. (67). The last line uses π^S being an isomorphism between T and X^S . The proof works similarly for Ω^M -local measurement noise. ■

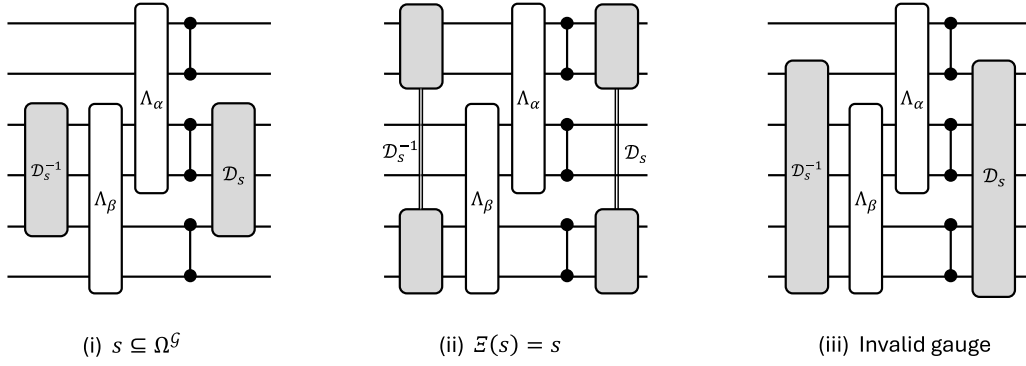


FIG. 6. Examples of three different cases of SDGs ($\mathcal{D}_s$) acting on $\mathcal{G} = \text{CZ}^{\otimes 3}$ with covariant Ω -local noise model, $\Omega^* = \{\{1, 2, 3, 4\}, \{3, 4, 5, 6\}\}$. Cases (i) and (ii) are allowed gauges as shown in Theorem 11 (b). Specifically, in case (i) we have $s = \{3, 4, 5\} \in \Omega$, thus $\mathcal{D}_s$ is a valid gauge that non-trivially transform the noise channel; In case (ii) we have $s = \{1, 2, 5, 6\} = \Xi_{\mathcal{G}}(s)$, or in words, any Pauli supports within s still supports within s after the action of $\mathcal{G}$ (see Lemma 10). Here, even though $s \notin \Omega$, $\mathcal{D}_s$ commutes through $\mathcal{G}$ leaving the noise channel unchanged, and is thus a trivially valid gauge; Case (iii) shows an invalid gauge ($s = \{2, 3, 4, 5, 6\}$), which would generate a 6-body noise term, violating the locality assumption.

Proof of Theorem 11 (b). Let $\pi^{\mathcal{G}} : X \mapsto X^{\mathcal{G}}$ be the projection map $\pi^{\mathcal{G}}(x) = x^{\mathcal{G}}$. Recall from Eq. (38) that for all $a \neq I_n, v \neq \emptyset$,

$$\begin{aligned} e_a^{\mathcal{G}} \cdot \mathfrak{d}_v &= \mathbb{1}[\text{pt}(\mathcal{G}(a))_v \neq 0_v] - \mathbb{1}[\text{pt}(a)_v \neq 0_v], \\ &= -\log(\langle \sigma_a | \mathcal{G}^\dagger \mathcal{D}_v \mathcal{G} \mathcal{D}_v^{-1} | \sigma_a \rangle). \end{aligned} \quad (73)$$

Here $\mathcal{D}_v$ is the subsystem depolarizing channel on v defined in Eq. (40).

- (1) If $\Xi_{\mathcal{G}}(v) = v$, then $\mathcal{G}$ is a tensor product of gates on v and on $[n] \setminus v$. Consequently, $\mathcal{G}$ commutes with $\mathcal{D}_v$. Thus, $e_a^{\mathcal{G}} \cdot \mathfrak{d}_v = 0$ for all a , which means $\pi^{\mathcal{G}}(\mathfrak{d}_v) \in \text{Im } \mathcal{Q}^{\mathcal{G}}$ trivially holds. See Fig. 6(ii) for an example.
- (2) If $v \in \Omega^{\mathcal{G}}$, one can see that both $\mathcal{D}_v$ and $\mathcal{D}_v^{-1}$ are $\Omega^{\mathcal{G}}$ -local Pauli channels. Since the noise model is $\mathcal{G}$ -covariant, $\mathcal{G}^\dagger \mathcal{D}_v \mathcal{G}$ is $\Omega^{\mathcal{G}}$ -local. Since the concatenation of two $\Omega^{\mathcal{G}}$ -local Pauli channels is clearly also $\Omega^{\mathcal{G}}$ -local, $\mathcal{G}^\dagger \mathcal{D}_v \mathcal{G} \mathcal{D}_v^{-1}$ is Ω -local. This means,

$$e_a^{\mathcal{G}} \cdot \mathfrak{d}_v = \sum_{b \prec a, b \sim \Omega^{\mathcal{G}}} r'_b, \quad (74)$$

for some coefficients $\{r'_b \in \mathbb{R} : b \sim \Omega^{\mathcal{G}}, b \neq I_n\}$. Comparing this to the definition of quasi-local gate noise, we see that $\pi^{\mathcal{G}}(\mathfrak{d}_v) \in \text{Im } \mathcal{Q}^{\mathcal{G}}$. See Fig. 6(i) for an example.

Combining the above two cases and using linearity, we complete the proof for Theorem 11 (b). $\blacksquare$

We make a few remarks about Theorem 11 (b). First, it is clear from the proof that $\text{span}\{\mathfrak{d}_v : \Xi_{\mathcal{G}}(v) = v\}$ are gauges that pass trivially through $\mathcal{G}$ leaving its noise parameters unchanged, while $\text{span}\{\mathfrak{d}_v : v \in \Omega^{\mathcal{G}}\}$ are gauges that

non-trivially transform the noise parameters of $\mathcal{G}$; Second, here we do not have a converse side as in Theorem 10 (b), meaning there could be more allowed gauges than what we have characterized. See Fig. 6 for an example of different types of gauges. We do not pursue such a tight characterization in the quasi-local case as it looks considerably more complicated. For our purpose of proving Theorem 11 (c), the current results are sufficient; Finally, the requirement of noise being $\mathcal{G}$ -covariant is indispensable for the theorem to hold. In Fig. 7 we show a counterexample where $\mathfrak{d}_v$ is not an allowed gauge even if $v \in \Omega^{\mathcal{G}}$, when the gate noise model is not $\mathcal{G}$ -covariant. This will be further discussed in Sec. VII when we encounter some realistic examples.

Proof of Theorem 11 (c). Thanks to Lemma 6, $\mathcal{Q}(T_R)$ is given by the intersection of Eq. (64) for SPAM and Eq. (65) for all $\mathcal{G} \in \mathfrak{G}$. Under the given conditions, it is easy to see that Eq. (64) is always a subset of Eq. (65) for any $\mathcal{G} \in \mathfrak{G}$, thus their intersection is simply given by Eq. (64). $\blacksquare$

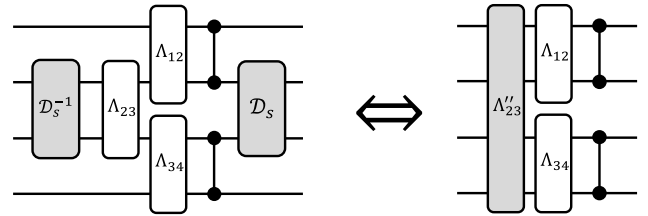


FIG. 7. Example of SDG acting on $\mathcal{G} = \text{CZ}^{\otimes 2}$ with nearest-neighbor two-local noise model (i.e., $\Omega^* = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$) that is *not* $\mathcal{G}$ -covariant. In this case, although $s = \{2, 3\} \in \Omega$, the SDG $\mathcal{D}_s$ would create a four-body noise term (shown on the right). This violates the Ω -local assumption and is therefore not an allowed gauge. This highlights the necessity of $\mathcal{G}$ -covariance in obtaining Theorem 11 (b).

C. Learning algorithm for quasi-local model

We now discuss algorithms for learning a quasi-local noise model. In this section, we will assume all conditions of Theorem 11 (c) are satisfied. That is, $\Omega^S \subseteq \Omega^M$; Ω^G is $\mathcal{G}$ -covariant for all $\mathcal{G} \in \mathfrak{G}$; and that $\forall \nu \in \Omega^S, \forall \mathcal{G} \in \mathfrak{G}$ either $\nu \in \Omega^G$ or $\Xi_{\mathcal{G}}(\nu) = \nu$. For such a well-conditioned quasi-local noise model, we can easily find a reduced cycle basis for L_R as follows.

Consider the following subset of the rooted cycle basis $\mathbf{B}$ defined in Eq. (18), denoted by $\mathbf{B}''$,

$$\mathbf{B}'' := \{\mathbf{e}_{\mu}^S + \mathbf{e}_{\mu}^M : \mu \in \Omega^M\} \cup \{\mathbf{e}_{\text{supp}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{supp}(G(a))}^M : \mathcal{G} \in \mathfrak{G}, a \sim \Omega^G\}. \quad (75)$$

Lemma 11. Let $\mathbf{B}_R := \mathcal{Q}^T(\mathbf{B}'')$. Then $\mathbf{B}_R$ forms a reduced cycle basis for L_R .

Proof. First, acting $\mathcal{Q}^T$ on the rooted cycle basis $\mathbf{B}$ defined in Lemma 5 yields,

$$\begin{aligned} \mathcal{Q}^T(\mathbf{e}_{\mu}^S + \mathbf{e}_{\mu}^M) &= \sum_{\substack{\nu \subseteq \mu, \\ \nu \sim \Omega^S}} \boldsymbol{\vartheta}_{\nu}^S + \sum_{\substack{\nu \subseteq \mu, \\ \nu \sim \Omega^M}} \boldsymbol{\vartheta}_{\nu}^M, & \forall \mu \neq \emptyset, \\ \mathcal{Q}^T(\mathbf{e}_{\text{supp}(a)}^S + \mathbf{e}_a^G + \mathbf{e}_{\text{supp}(G(a))}^M) &= \sum_{\substack{\nu \subseteq \text{supp}(a), \\ \nu \sim \Omega^S}} \boldsymbol{\vartheta}_{\nu}^S + \sum_{\substack{b \sim a, \\ b \sim \Omega^G}} \boldsymbol{\vartheta}_b^G + \sum_{\substack{\nu \subseteq \text{supp}(G(a)), \\ \nu \sim \Omega^M}} \boldsymbol{\vartheta}_{\nu}^M, & \forall a \neq I_n, \quad \forall \mathcal{G} \in \mathfrak{G}. \end{aligned} \quad (76)$$

Thus, $\mathbf{B}_R := \mathcal{Q}^T(\mathbf{B}'')$ takes the following form:

$$\mathbf{B}_R = \left\{ \sum_{\substack{\nu \subseteq \mu, \\ \nu \sim \Omega^S}} \boldsymbol{\vartheta}_{\nu}^S + \sum_{\nu \subseteq \mu} \boldsymbol{\vartheta}_{\nu}^M : \mu \in \Omega^M \right\} \cup \left\{ \sum_{\substack{\nu \subseteq \text{supp}(a), \\ \nu \sim \Omega^S}} \boldsymbol{\vartheta}_{\nu}^S + \sum_{b \sim a} \boldsymbol{\vartheta}_b^G + \sum_{\substack{\nu \subseteq \text{supp}(G(a)), \\ \nu \sim \Omega^M}} \boldsymbol{\vartheta}_{\nu}^M : \mathcal{G} \in \mathfrak{G}, a \sim \Omega^G \right\}, \quad (77)$$

where we have used the defining property of Ω^M and Ω^G . let us first prove the linear independence of $\mathbf{B}_R$. Consider a linear combination of all vectors from $\mathbf{B}_R$ that yields $\mathbf{0}$. Since only the second part of $\mathbf{B}_R$ contains gate noise parameters, and one can show $\{\sum_{b \sim a} \boldsymbol{\vartheta}_b^G : \mathcal{G} \in \mathfrak{G}, a \sim \Omega^G\}$ are linearly-independent (as they can represent $\{\boldsymbol{\vartheta}_a^G : \mathcal{G} \in \mathfrak{G}, a \sim \Omega^G\}$ using the inclusion-exclusion principle, see Appendix E for similar expressions), the coefficients for all vectors from the second part must be zero; On the other hand, because $\{\sum_{\nu \subseteq \mu} \boldsymbol{\vartheta}_{\nu}^M : \mu \in \Omega^M\}$ are also linearly-independent (also thanks to the inclusion-exclusion principle), the coefficients for vectors from the first part must also be zero. This means $\mathbf{B}_R$ must be linearly independent. On the other hand, Theorem 11 (c) implies $\dim(T_R) = |\Omega^S|$. We can see from the expression of $\mathbf{B}_R$ that $|\mathbf{B}_R| = \dim(X_R) - |\Omega^S| = \dim(X_R) - \dim(T_R) = \dim(L_R)$. We can thus conclude that $\mathbf{B}_R$ forms a basis for L_R . ■

The construction of $\mathbf{B}_R$ implies the following set of experiments can fully learn the quasi-local reduced noise model

- (1) For all $\mu \in \Omega^M$, prepare $\tilde{\rho}_0$ and measure the expectation value of $\tilde{Z}^{\mu}$.

- (2) For all $\mathcal{G} \in \mathfrak{G}$, $b \sim \Omega^G$, prepare $\tilde{\rho}_b$, apply $\tilde{\mathcal{G}}$, and measure $\tilde{P}_{\mathcal{G}(b)}$.

If Ω^M and Ω^G have polynomial size, the number of experiments needed is also polynomial, and is thus efficient. Besides, if each $\mathcal{G}$ is a parallel application of many Clifford gates each acting on a constant number of qubits, many of the above experiments can be conducted in parallel. We leave such an optimization of experiment design as future research.

Of course, one might also want to learn gate noise parameter to relative precision in the quasi-local noise model, which is related to finding a reduced cycle basis for L_R^G . We will not pursue a general efficient algorithm for this purpose due to the complexity of the quasi-local model. In fact, we conjecture that it is computationally hard to establish such a basis in general. Instead of dealing with the general cases, we will study one practically-interesting examples in the following section.

VII. CASE STUDY: CZ GATES WITH 1D TOPOLOGY

In this section, we will apply our theory to a concrete gate set where the CZ gates are the only multi-qubit entangling gates. Such a gate set, augmented with single-qubit

TABLE II. Summary of results of this section. Each row corresponds to a gate set and noise model studied in one subsection.

Section	n	$\mathcal{G}$	Noise model	$\dim(X_R)$	$\dim(L_R)$	$\dim(X_R^G)$	$\dim(L_R^G)$
VII A	2	{CZ}	complete	21	18	15	13
VII B	$n \geq 4$	{CZ _{<i>i,i+1</i>} }	fully local	$17n$	$16n$	$15n$	$14n$
VII C	$n \geq 6$	{CZ _{<i>e</i>} , CZ _{<i>o</i>} }	nearest-neighbor	$28n$	$27n$	$24n$	$23n$
VII D	$n \geq 6$	{CZ _{<i>e</i>} , CZ _{<i>o</i>} }	covariant 4-local	$244n$	$242n$	$240n$	—

gates, are able to conduct universal quantum computation, and is adopted by state-of-the-art experimental platforms (e.g., IBM Quantum [37] [81]). We will discuss the noise learnability and learning algorithms with several different noise ansatz. We hope the results presented here can not only illustrate our theoretical results, but also serve as a new noise characterization protocol for practical use. A quick summary of results is given in Table II.

A. Single CZ, complete noise

We begin with a minimal example: $n = 2$, $\mathcal{G} = \{\text{CZ}\}$. The goal is to learn the complete Pauli noise model. Recall that the associated PTG for this model is given in Fig. 2. Thanks to Theorem 7, the parameter space X , gauge space T , and learnable space L are given by (let $\mathcal{G} = \text{CZ}$ for notational simplicity),

- (1) X : $\dim(X) = 21$. Basis: $\{e_u^S, e_u^M, e_a^G : u \neq 00, a \neq II\}$.
- (2) T : $\dim(T) = 3$. Basis: $\{\mathfrak{d}_{10}, \mathfrak{d}_{01}, \mathfrak{d}_{11}\}$.
- (3) L : $\dim(L) = 18$. Cycle basis:

$$\begin{aligned} & \{e_u^S + e_u^M : u \in \{01, 10, 11\}\} \\ & \cup \{e_{\text{pt}(a)}^S + e_a^G + e_{\text{pt}(\mathcal{G}(a))}^M : a \in \mathbb{P}^2 \setminus \{II\}\}. \end{aligned}$$

The learnable space for gate noise $L^G := L \cap X^G$ (which is the cycle space for the subgraph of PTG with edges only for gate noise parameters) is spanned by the following cycle basis:

$$\begin{aligned} & \{e_{IZ}^G, e_{ZI}^G, e_{ZZ}^G, e_{XX}^G, e_{YY}^G, e_{XY}^G, e_{YX}^G, \\ & e_{XI}^G + e_{XZ}^G, e_{YI}^G + e_{YZ}^G, e_{XI}^G + e_{YZ}^G, \\ & e_{IX}^G + e_{ZX}^G, e_{IY}^G + e_{ZY}^G, e_{IX}^G + e_{ZY}^G\}. \end{aligned} \quad (78)$$

We have $\dim(L^G) = 13$. Note that $\dim(X^G) - \dim(L^G) = 2$. This is because $\pi^G(\mathfrak{d}_{11}) = \mathbf{0}$, which means the global depolarizing gauge does not change gate noise parameters. Only the two single-qubit depolarizing gauges $\{\mathfrak{d}_{01}, \mathfrak{d}_{10}\}$ change gate parameters non-trivially. The above basis for L^G can be augmented to a basis of L by adding the following five rooted cycles:

$$\begin{aligned} & \{e_u^S + e_u^M : u \in \{01, 10, 11\}\} \\ & \cup \{e_{10}^S + e_{XI}^G + e_{11}^M, e_{01}^S + e_{IX}^G + e_{11}^M\}. \end{aligned} \quad (79)$$

We remark that Eq. (78) was previously given in Ref. [18]. The new insight from our theory is to include SPAM parameters as the target of learning, and to explicitly parameterize the gauge degrees of freedom. We also note that Ref. [18] has presented experimental designs to learn all of these gate parameters to relative precision. Concretely, any of the above parameters can be learned by concatenating CZ or CZ $\circ (\sqrt{Z} \otimes \sqrt{Z})$ an even number of time, input the correct Pauli eigenstate, and measurement in the correct Pauli eigenbasis. See Fig. 8(a).

B. CZs on ring, fully local noise

Let $n > 2$, consider the case where the n qubits are placed in a 1D ring, and one can apply an CZ between any neighboring pair of qubits, i.e., $\mathcal{G} = \{\text{CZ}_{i,i+1} : i \in [n]\}$, where the $(n+1)$ -th qubit is the same as the first qubit. Suppose the noise model is fully local (cf. Definition 18), using Theorem 10, we obtain the following characterization for the learnability of this model (let $\mathcal{G}_i = \text{CZ}_{i,i+1}$),

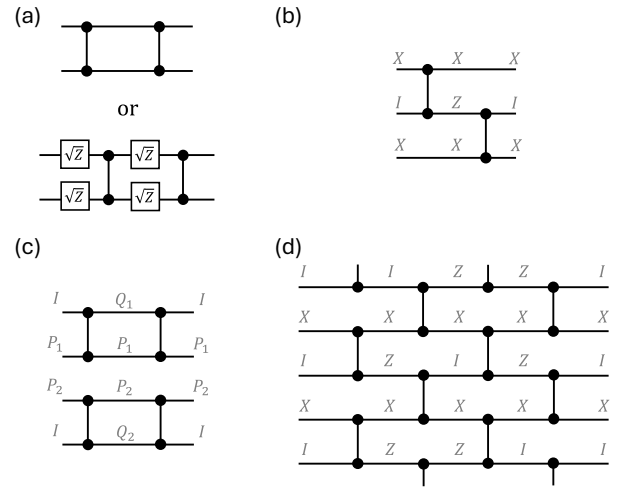


FIG. 8. Circuits for learning different types of cycles. (a) Circuits for learning all cycles of a single CZ (and for certain cycles of the other models). Any cycles from Eq. (78) can be learned by inputting an appropriate Pauli eigenstate to the upper or lower circuit and measure the same Pauli observable. (b) Circuits for learning the two-gate cycle from Eq. (81). The evolution of the corresponding Pauli operators is shown in gray. (c) Circuits for learning cycles from Eq. (85). (d) Circuits for learning for cycles from Eq. (86).

- (1) X_R : $\dim(X_R) = 17n$. Basis: $\{\vartheta_j^S, \vartheta_j^M, \vartheta_a^{G_i} : j \in [n], i \in [n], a \in \mathbb{P}^2 \setminus II\}$.
- (2) T_R : $\dim(T_R) = n$. Basis: $\{\mathcal{Q}^{-1}(\mathfrak{d}_{\{i\}}) : i \in [n]\}$.
- (3) L_R : $\dim(L_R) = 16n$. Corresponding cycle basis: $\{e_{\{i\}}^S + e_{\{i\}}^M : i \in [n]\} \cup \{e_{\text{pt}(a)}^S + e_a^G + e_{\text{pt}(G_i(a))}^M : i \in [n], a \in \mathbb{P}^{[i,i+1]} \setminus I_n\}$.

Note that $\mathcal{Q}^{-1}$ is well-defined when acting within the image of $\mathcal{Q}$. For L_R , instead of writing down a reduced cycle basis, we write down the corresponding cycle basis in L whose connection to experiment design is more direct. Acting $\mathcal{Q}^\top$ on that basis will give a reduced cycle basis for L_R .

Using the algorithms introduced in Appendix D., we can construct a reduced cycle basis for the learnable spaces of gate parameters L_R^G . In fact, there exists a basis where each basis vector contains parameters of only one gate or two neighboring gates, as follows:

- (1) One-gate [same as Eq. (78)]:

$$\begin{aligned} &\{\vartheta_{IZ}^{G_i}, \vartheta_{ZI}^{G_i}, \vartheta_{ZZ}^{G_i}, \vartheta_{XX}^{G_i}, \vartheta_{YY}^{G_i}, \vartheta_{XY}^{G_i}, \vartheta_{YX}^{G_i}, \\ &\quad \vartheta_{XI}^{G_i} + \vartheta_{XZ}^{G_i}, \vartheta_{YI}^{G_i} + \vartheta_{YZ}^{G_i}, \vartheta_{XI}^{G_i} + \vartheta_{YZ}^{G_i}, \vartheta_{IX}^{G_i} + \vartheta_{ZX}^{G_i}, \\ &\quad \vartheta_{IY}^{G_i} + \vartheta_{ZY}^{G_i}, \vartheta_{IX}^{G_i} + \vartheta_{ZY}^{G_i} : i \in [n]\}. \end{aligned} \quad (80)$$

- (2) Two-gate:

$$\{\vartheta_{XI}^{G_i} + \vartheta_{ZX}^{G_{i+1}} : i \in [n]\}. \quad (81)$$

We have $\dim(L^G) = 14n$. The one-gate basis vector can be learned by running the same learning protocol for a single CZ in each consecutive two-qubit subsystem, shown in Fig. 8(a); The two-gate basis can be learned using the circuits shown in Fig. 8(b) on each consecutive three-qubit subsystem. By learning disjoint regions in parallel, a constant number of experiments is sufficient to learn all the gate noise parameters. We also note that the above basis for L_R^G can be augmented to a basis for L_R by adding the following $2n$ reduced cycles,

$$\begin{aligned} &\{\vartheta_j^S + \vartheta_j^M : j \in [n]\} \\ &\cup \{\vartheta_j^S + \vartheta_{XI}^{G_j} + \vartheta_j^M + \vartheta_{j+1}^M : j \in [n]\}. \end{aligned} \quad (82)$$

C. CZs on ring, nearest-neighbor noise

Let $n \geq 6$ be an even number. We again consider n qubits placed in an 1D ring, but this time let the gate set be $\mathfrak{G} = \{\mathcal{G}_e, \mathcal{G}_o\}$ where

$$\begin{aligned} \mathcal{G}_e &= \text{CZ}_{1,2} \otimes \text{CZ}_{3,4} \otimes \cdots \otimes \text{CZ}_{n-1,n}, \\ \mathcal{G}_o &= \text{CZ}_{2,3} \otimes \text{CZ}_{4,5} \otimes \cdots \otimes \text{CZ}_{n,1}, \end{aligned} \quad (83)$$

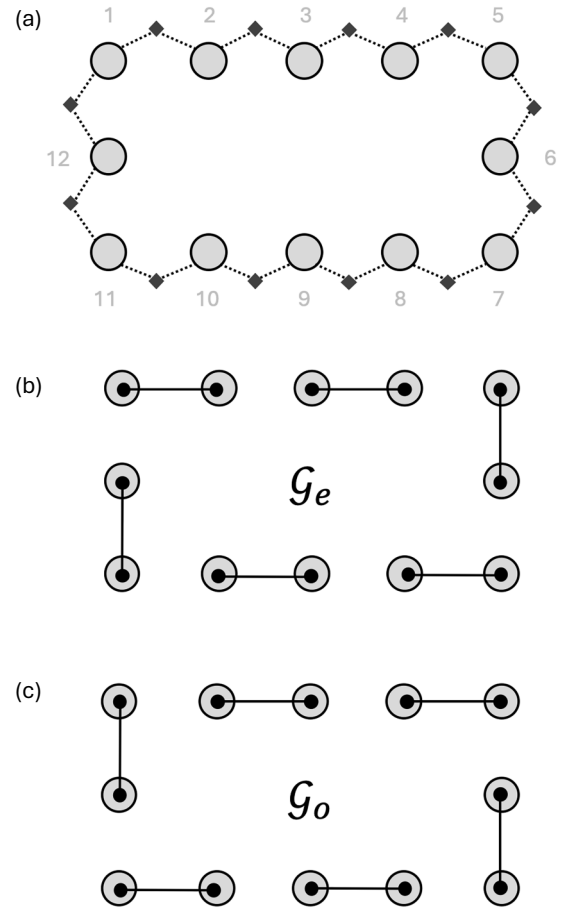


FIG. 9. Illustration of the setup in Sec. VII C with $n = 12$. (a) Qubit topology and noise model. Each circle represents a qubit with their label annotated. Each diamond connects to a maximal factor from Ω_* , specifying the nearest-neighbor noise model for both SPAM and gate noises. (b),(c) Illustration of $\mathcal{G}_e, \mathcal{G}_o$, respectively.

and we assume the state preparation, measurement, and gate noise for both $\mathcal{G}_e, \mathcal{G}_o$ are all Ω -local, with Ω associated with $\Omega_* = \{\{i, i+1\} : i \in [n]\}$. The setup is illustrated in Fig. 9. This setup is essentially the one used in Ref. [49] for conducting an error-mitigated quantum simulation task. It is thus of great interest to understand the learnability of this noise model.

To begin with, this noise model is not gate-covariant in the sense of Definition 22. To see this, consider the Pauli channel $\Lambda'(\rho) = \eta X_2 X_3 \rho X_2 X_3 + (1 - \eta)\rho$ with some $\eta < 1/2$, which is Ω -local thanks to Lemma E2. We have $\mathcal{G}_e \Lambda' \mathcal{G}_e^\dagger = \eta Z_1 X_2 X_3 Z_4 \rho Z_1 X_2 X_3 Z_4 + (1 - \eta)\rho$, which is not Ω -local, again by Lemma E2. Similarly for $\mathcal{G}_o$. This means we cannot directly apply Theorem 11 (c). In fact, we can show that the reduced gauge space is given by $T_R = \text{span}\{\mathcal{Q}^{-1}(\mathfrak{d}_{\{i\}}) : i \in [n]\}$, i.e., only the single-qubit depolarizing gauges. Intuitively, any (linear combination of) two-qubit depolarizing gauge transformation will violate

the Ω -local noise assumption of either $\mathcal{G}_e$ or $\mathcal{G}_o$. A formal proof is given in Appendix F 1. Therefore, we obtain the following characterization of the model learnability:

- (1) X_R : $\dim(X_R) = 28n$. Basis: $\{\vartheta_v^S, \vartheta_v^M, \vartheta_a^{\mathcal{G}_e}, \vartheta_a^{\mathcal{G}_o} : v \in \Omega, a \sim \Omega\}$.
- (2) T_R : $\dim(T_R) = n$. Basis: $\{\mathcal{Q}^{-1}(\mathfrak{d}_{\{i\}}) : i \in [n]\}$.
- (3) L_R : $\dim(L_R) = 27n$.

Note that the cardinality of $\{a \in \mathbf{P}^n \setminus I_n : a \sim \Omega\}$ can be computed using the inclusion-exclusion principle: $(4^2 - 1)n - (4 - 1)n = 12n$.

We have also found a reduced cycle basis for L_R^G . To make the connection to experiments more straightforward, we will write down a cycle basis that generates the reduced cycle basis by applying $\mathcal{Q}^T$ on them (cf. Definition 15). The cycle bases are divided into two parts, depending on whether the basis vector involves one or two gates, as follows:

- (1) One-gate:

- (a) For $k = 1, \dots, \frac{n}{2}$, similar as Eq. (78),

$$\begin{aligned} & \{e_{I_{2k-1}Z_{2k}}^{\mathcal{G}_e}, \dots, e_{X_{2k-1}I_{2k}}^{\mathcal{G}_e} + e_{X_{2k-1}Z_{2k}}^{\mathcal{G}_e}, \dots\} \\ & \cup \{e_{I_{2k}Z_{2k+1}}^{\mathcal{G}_o}, \dots, e_{X_{2k}I_{2k+1}}^{\mathcal{G}_o} + e_{X_{2k}Z_{2k+1}}^{\mathcal{G}_o}, \dots\}. \end{aligned} \quad (84)$$

- (b) For $k = 1, \dots, \frac{n}{2}$,

$$\begin{aligned} & \{e_a^{\mathcal{G}_e} + e_{a'}^{\mathcal{G}_e} : a \in \mathbf{P}^{[2k, 2k+1]}, w(a) = 2, a' = \mathcal{G}_e(a) \\ & \cup \{e_a^{\mathcal{G}_o} + e_{a'}^{\mathcal{G}_o} : a \in \mathbf{P}^{[2k+1, 2k+2]}, w(a) = 2, \\ & a' = \mathcal{G}_o(a)\}. \end{aligned} \quad (85)$$

- (2) Two-gate: For $k = 1, \dots, \frac{n}{2}$,

$$\begin{aligned} & \{e_{IXIXI_{[2k-1, \dots, 2k+3]}}^{\mathcal{G}_o} + e_{IXXZX_{[2k-1, \dots, 2k+3]}}^{\mathcal{G}_e} \\ & + e_{ZXIXZ_{[2k-1, \dots, 2k+3]}}^{\mathcal{G}_o} + e_{ZXZXI_{[2k-1, \dots, 2k+3]}}^{\mathcal{G}_e}\} \\ & \cup \{e_{IXIXI_{[2k, \dots, 2k+4]}}^{\mathcal{G}_o} + e_{ZXZXI_{[2k, \dots, 2k+4]}}^{\mathcal{G}_e} \\ & + e_{ZXIXZ_{[2k, \dots, 2k+4]}}^{\mathcal{G}_o} + e_{IXXZX_{[2k, \dots, 2k+4]}}^{\mathcal{G}_e}\}. \end{aligned} \quad (86)$$

Equations (84), (85), (86) contribute $13n$, $9n$, n cycles, respectively, and is consistent with the fact that $\dim(L_R^G) = 23n$. Elusive as they might look, each type of cycle corresponds to an explicit family of experiment designs, as shown in Figs. 8(a), 8(c), 8(d), respectively.

Finally, we note that a constant number of circuits (each with sufficiently many shots) suffice to learn this reduced cycle basis. This is because the circuits shown in Figs. 8(a), 8(c), 8(d) only affect a constant number of qubits and can

thus be run in parallel. While we leave a more detailed analysis of the sample complexity and the minimal number of circuits for future studies, we provide a notebook illustrating how to design learning circuits for the noise model discussed in this section [39].

For a proof that the above cycles constitute a cycle basis for L_R^G [82], and a way to supplement it into a complete basis for L_R , see Appendix F 2.

D. CZs on ring, covariant quasi-local noise

Finally, we present a minimal example of quasi-local noise model that is gate-covariant. Let $n \geq 6$ be even. Consider the same gate set $\mathfrak{G} = \{\mathcal{G}_e, \mathcal{G}_o\}$ as in the last section. Assume the SPAM noise is still Ω -local for $\Omega_* = \{\{i, i+1\} : i \in [n]\}$. Assume the noise for $\mathcal{G}_e, \mathcal{G}_o$ is $\Omega^{\mathcal{G}_e}, \Omega^{\mathcal{G}_o}$ -local, respectively, where

$$\begin{aligned} \Omega_*^{\mathcal{G}_e} &= \{\{2k-1, 2k, 2k+1, 2k+2\} : k \in [n/2]\}, \\ \Omega_*^{\mathcal{G}_o} &= \{\{2k, 2k+1, 2k+2, 2k+3\} : k \in [n/2]\}. \end{aligned} \quad (87)$$

The factor graphs for each noise model are given in Fig. 10. The fact that this model is gate-covariant follows from Lemma 10. Indeed, each maximal factor is weight-4, containing exactly the support of two neighboring CZ. Consequently, Theorem 11 gives us that

- (1) X_R : $\dim(X_R) = 244n$. Basis: $\{\vartheta_v^S, \vartheta_v^M, \vartheta_a^{\mathcal{G}_e}, \vartheta_b^{\mathcal{G}_o} : v \in \Omega, a \sim \Omega^{\mathcal{G}_e}, b \sim \Omega^{\mathcal{G}_o}\}$.
- (2) T_R : $\dim(T_R) = 2n$. Basis: $\{\mathcal{Q}^{-1}(\mathfrak{d}_v) : v \in \Omega\}$.

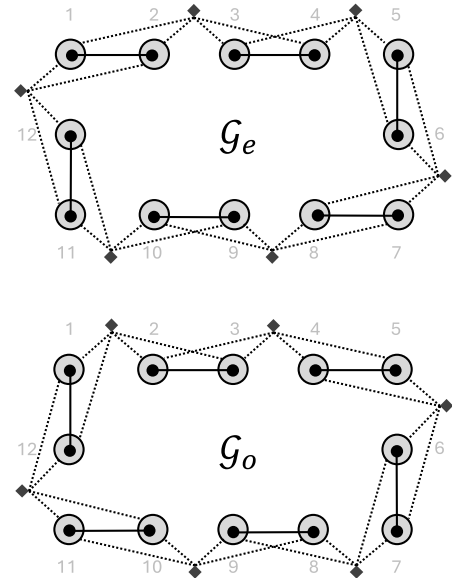


FIG. 10. The covariant quasi-local gate noise model from Sec. VII D. The left and right shows Ω^e, Ω^o , respectively, with each diamond connecting to a maximal factor from Ω_*^e (respectively, Ω_*^o).

- (3) L_R : $\dim(L_R) = 242n$. Corresponding cycle basis:
 $\{e_v^S + e_v^M : v \in \Omega\} \cup \{e_{\text{pt}(a)}^S + e_a^G + e_{\text{pt}(\mathcal{G}(a))}^M : \mathcal{G} \in \{\mathcal{G}_e, \mathcal{G}_o\}, a \sim \Omega^G\}$.

Note that the cardinality of $\{a \in \mathbf{P}^n \setminus I_n : a \sim \Omega^{G_e}\}$ can again be computed using the inclusion-exclusion principle: $\frac{n}{2} \times (4^4 - 4^2) = 120n$. One can see that a linear number of experiments that involves measuring at most weight-4 Pauli observables is sufficient to learn the whole noise model. It is also not hard to see that, by paralleling many of the experiments, a constant number of experiments would also suffice.

We will not pursue a reduced cycle basis for L_R^G in this example due to its complexity, and leave this for future investigation.

Notes added. We became aware of an independent and concurrent work [83] that also studied the learnability of the sparse Pauli-Lindblad noise model and proposed novel learning protocols. We thank Alessio Calzona for communicating with us about their manuscript.

ACKNOWLEDGMENTS

We would like to thank Alireza Seif, Yunchao Liu, Edward Chen, Joseph Emerson, Matthew Graydon, Hsin-Yuan Huang, Ming Yuan, Sisi Zhou, and Yunxiao Zhao for helpful discussions. S.C. and L.J. acknowledge support from the ARO (W911NF-23-1-0077), ARO MURI (W911NF-21-1-0325), AFOSR MURI (FA9550-21-1-0209, FA9550-23-1-0338), DARPA (HR0011-24-9-0359, HR0011-24-9-0361), NSF (ERC-1941583, OMA-2137642, OSI-2326767, CCF-2312755, OSI-2426975), and Packard Foundation (2020-71479).

DATA AVAILABILITY

The data that support the findings of this article are openly available [84].

APPENDIX A: PROOF OF THEOREM 6

The proof follows [18, Appendix D] with some modification to fit in our framework.

Proof of Theorem 6 (a). For any experiment F , we need to find a finite set of rooted cycles (i.e., linear functions corresponding to a rooted cycle) $\{f_j\}_{j=1}^M$ and a function $\hat{F}$ such that

$$F(\mathbf{x}) = \hat{F}(f_1(\mathbf{x}), \dots, f_M(\mathbf{x})), \quad \forall \mathbf{x} \in X. \quad (\text{A1})$$

By definition, $F(\mathbf{x})$ is a 2^n -dimensional real vector such that $F_k(\mathbf{x}) = \text{Tr}[\tilde{E}_k \tilde{\mathcal{C}}(\tilde{\rho}_0)]$ where the noisy initial state, circuit, and POVM element are determined by the noise parameter $\mathbf{x}$. Expanding the SPAM noise parameters yields

$$F_k(\mathbf{x}) = \frac{1}{4^n} \sum_{u, u' \in \{0,1\}^n} (-1)^{k \cdot u'} \lambda_{u'}^M \lambda_u^S \text{Tr}[Z^{u'} \tilde{\mathcal{C}}(Z^u)]. \quad (\text{A2})$$

On the other hand, any noisy circuit $\mathcal{C}$ satisfying our assumptions can be written as

$$\tilde{\mathcal{C}} = \mathcal{U}^{(m)} \circ \tilde{\mathcal{G}}_m \circ \dots \circ \mathcal{U}^{(1)} \circ \tilde{\mathcal{G}}_1 \circ \mathcal{U}^{(0)}, \quad (\text{A3})$$

where each $\mathcal{U}^{(i)} = \bigotimes_{j=1}^n \mathcal{U}_j^{(i)}$ is a layer of parallel single-qubit gate, and each $\tilde{\mathcal{G}}_i$ is the noisy realization of a Clifford gate $\mathcal{G}_i$ from our gate set $\mathcal{G}$. Note that consecutive layers of single-qubit gates can always be merged into a single layer, and we have assumed single-qubit gate to be noiseless, so the above form of $\tilde{\mathcal{C}}$ is indeed general.

An important property for single-qubit gate layers is that, even if they are non-Clifford, they always preserve the pattern of an input Pauli operator. That is,

$$\mathcal{U}^{(i)}(P_a) = \sum_{b: \text{pt}(b)=\text{pt}(a)} c_{b,a}^{(i)} P_b, \quad \forall P_a \in \mathbf{P}^n, \quad (\text{A4})$$

for some real coefficients $c_{b,a}^{(i)}$. This comes from the fact that any single-qubit unitary channel is trace-preserving and unital. Thanks to this property, the action of $\tilde{\mathcal{C}}$ on any input Pauli operator can be expressed as

$$\begin{aligned} \tilde{\mathcal{C}}(P_a) &= (\mathcal{U}^{(m)} \circ \tilde{\mathcal{G}}_m \circ \dots \circ \mathcal{U}^{(1)} \circ \tilde{\mathcal{G}}_1 \circ \mathcal{U}^{(0)})(P_a) \\ &= (\mathcal{U}^{(m)} \circ \tilde{\mathcal{G}}_m \circ \dots \circ \mathcal{U}^{(1)} \circ \tilde{\mathcal{G}}_1) \left(\sum_{b_0: \text{pt}(b_0)=\text{pt}(a)} c_{b_0,a}^{(0)} P_{b_0} \right) \\ &= (\mathcal{U}^{(m)} \circ \tilde{\mathcal{G}}_m \circ \dots \circ \mathcal{U}^{(1)}) \left(\sum_{b_0: \text{pt}(b_0)=\text{pt}(a)} c_{b_0,a}^{(0)} \lambda_{b_0}^{\mathcal{G}_1} P_{\mathcal{G}_1(b_0)} \right) \end{aligned}$$

$$\begin{aligned}
&= (\mathcal{U}^{(m)} \circ \tilde{\mathcal{G}}_m \circ \dots \circ \mathcal{U}^{(2)}) \left(\sum_{\substack{b_0: \text{pt}(b_0)=\text{pt}(a), \\ b_1: \text{pt}(b_1)=\text{pt}(\mathcal{G}_1(b_0))}} c_{b_1, \mathcal{G}_1(b_0)}^{(1)} c_{b_0, a}^{(0)} \lambda_{b_1}^{\mathcal{G}_2} \lambda_{b_0}^{\mathcal{G}_1} P_{\mathcal{G}_2(b_1)} \right) \\
&= \dots \\
&= \sum_{\substack{b_0: \text{pt}(b_0)=\text{pt}(a), \\ b_1: \text{pt}(b_1)=\text{pt}(\mathcal{G}_1(b_0)), \\ \vdots \\ b_m: \text{pt}(b_m)=\text{pt}(\mathcal{G}_m(b_{m-1}))}} c_{b_m, \mathcal{G}_m(b_{m-1})}^{(m)} \dots c_{b_1, \mathcal{G}_1(b_0)}^{(1)} c_{b_0, a}^{(0)} \lambda_{b_{m-1}}^{\mathcal{G}_m} \dots \lambda_{b_1}^{\mathcal{G}_2} \lambda_{b_0}^{\mathcal{G}_1} P_{b_m}. \tag{A5}
\end{aligned}$$

Now, substitute this back into the expression of $F_k(\mathbf{x})$,

$$\begin{aligned}
F_k(\mathbf{x}) &= \sum_{\substack{u, u' \in \{0,1\}^n, \\ b_0: \text{pt}(b_0)=\text{pt}(u), \\ b_1: \text{pt}(b_1)=\text{pt}(\mathcal{G}_1(b_0)), \\ \vdots \\ b_m: \text{pt}(b_m)=\text{pt}(\mathcal{G}_m(b_{m-1}))}} \frac{(-1)^{k \cdot u'}}{4^n} c_{b_m, \mathcal{G}_m(b_{m-1})}^{(m)} \dots c_{b_1, \mathcal{G}_1(b_0)}^{(1)} c_{b_0, a}^{(0)} \lambda_{u'}^M \lambda_{b_{m-1}}^{\mathcal{G}_m} \dots \lambda_{b_1}^{\mathcal{G}_2} \lambda_{b_0}^{\mathcal{G}_1} \lambda_u^S \text{Tr}[Z^{u'} P_{b_m}] \\
&= \sum_{\substack{u \in \{0,1\}^n, \\ b_0: \text{pt}(b_0)=\text{pt}(u), \\ b_1: \text{pt}(b_1)=\text{pt}(\mathcal{G}_1(b_0)), \\ \vdots \\ b_m: \text{pt}(b_m)=\text{pt}(\mathcal{G}_m(b_{m-1}))}} \frac{(-1)^{k \cdot \text{pt}(b_m)}}{2^n} c_{b_m, \mathcal{G}_m(b_{m-1})}^{(m)} \dots c_{b_0, a}^{(0)} \lambda_{\text{pt}(b_m)}^M \lambda_{b_{m-1}}^{\mathcal{G}_m} \dots \lambda_{b_0}^{\mathcal{G}_1} \lambda_u^S \mathbb{1}[Z^{\text{pt}(b_m)} = P_{b_m}]. \tag{A6}
\end{aligned}$$

Let us look at the indexes of the summation. For $u = \mathbf{0}$, we must have $b_m = \dots = b_0 = I$, and the corresponding term in the sum becomes $1/2^n$ that is a constant; For $u \neq \mathbf{0}$, we must have $b_i \neq I$ for all i , and the corresponding term in the sum depends on the noise parameters via

$$\lambda_{\text{pt}(b_m)}^M \lambda_{b_{m-1}}^{\mathcal{G}_m} \dots \lambda_{b_1}^{\mathcal{G}_2} \lambda_{b_0}^{\mathcal{G}_1} \lambda_u^S = \exp\left(-\left(x_{\text{pt}(b_m)}^M + x_{b_{m-1}}^{\mathcal{G}_m} + \dots + x_{b_0}^{\mathcal{G}_1} + x_u^S\right)\right) =: \exp(-f_{b,u}(\mathbf{x})), \tag{A7}$$

where $f_{b,u}$ is a linear function of $\mathbf{x}$. More importantly, thanks to the restriction of $(b_0, \dots, b_m)$, one can see that $f_{b,u}$ is actually a rooted cycle in the pattern transfer graph. Therefore, $F_k(\mathbf{x})$ [and thus $F(\mathbf{x})$] is indeed determined by $\{f_{b,u}(\mathbf{x})\}_{b,u}$ consisting of finitely many rooted cycles. This completes the proof of Theorem 6 (a). ■

The following proof again comes from Ref. [18] with minor modification, and is presented here for completeness.

Proof of Theorem 6 (b). For any rooted cycles f , we need to find an experiment F and a function $\hat{f}$ such that

$$f(\mathbf{x}) = \hat{f}(F(\mathbf{x})), \quad \forall \mathbf{x} \in X. \tag{A8}$$

A generic rooted cycle f of depth m takes the following form (here, we write down the vertices for clarity, though

they are not needed to define a cycle),

$$\begin{aligned}
&(v_R, x_{b_0}^S, v_{b_0}, x_{P_1}^{\mathcal{G}_1}, v_{b_1}, x_{P_2}^{\mathcal{G}_2}, \dots, x_{P_m}^{\mathcal{G}_m}, v_{b_m}, x_{b_m}^M, v_R), \\
&\text{s.t. } b_{i-1} = \text{pt}(P_i), \text{pt}(\mathcal{G}_i(P_i)) = b_i, \quad \forall i = 1, \dots, m. \tag{A9}
\end{aligned}$$

Consider the following noisy circuit:

$$\tilde{\mathcal{C}} = \mathcal{U}^{(m)} \circ \tilde{\mathcal{G}}_m \circ \dots \circ \mathcal{U}^{(1)} \circ \tilde{\mathcal{G}}_1 \circ \mathcal{U}^{(0)},$$

where $\mathcal{U}^{(i)} \equiv \bigotimes_{j=1}^n \mathcal{U}_j^{(i)}$ is Clifford, and satisfies

$$\begin{aligned}
&\times \begin{cases} \mathcal{U}^{(0)}(Z^{b_0}) = P_1; \\ \mathcal{U}^{(i)}(\mathcal{G}_i(P_i)) = P_{i+1}, \quad \forall i = 1, \dots, m-1; \\ \mathcal{U}^{(m)}(\mathcal{G}_m(P_m)) = Z^{b_m}. \end{cases} \tag{A10}
\end{aligned}$$

Note that, thanks to the constraints from Eq. (A9), the above requirements for $\mathcal{U}^{(i)}$ are literally mapping one Pauli

operator to another *with the same pattern*, and thus can indeed be achieved by a layer of single-qubit Clifford gates.

This circuit yields the following experiment (i.e., experimental outcome distribution)

$$\begin{aligned} F_k(\mathbf{x}) &= \text{Tr}[\tilde{E}_k \tilde{\mathcal{C}}(\tilde{\rho}_0)] \\ &= \frac{1}{4^n} \sum_{u, u' \in \{0,1\}^n} (-1)^{k \cdot u'} \lambda_{u'}^M \lambda_u^S \text{Tr}[Z^{u'} \tilde{\mathcal{C}}(Z^u)], \\ &\quad \forall k \in \{0,1\}^n. \end{aligned} \quad (\text{A11})$$

Compute the following function of experimental outcome (which is, intuitively, the expectation value of Z^{b_m} on the equivalent noisy state right before a perfect measurement)

$$\begin{aligned} &\sum_{k \in \{0,1\}^n} (-1)^{k \cdot b_m} F_k(\mathbf{x}) \\ &= \frac{1}{4^n} \sum_{k, u, u'} (-1)^{k \cdot (b_m + u')} \lambda_{u'}^M \lambda_u^S \text{Tr}[Z^{u'} \tilde{\mathcal{C}}(Z^u)] \\ &= \frac{1}{2^n} \sum_u \lambda_{b_m}^M \lambda_u^S \text{Tr}[Z^{b_m} \tilde{\mathcal{C}}(Z^u)] \\ &= \frac{1}{2^n} \sum_u \lambda_{b_m}^M \lambda_{P_m}^{G_m} \dots \lambda_{P_1}^{G_1} \lambda_u^S \text{Tr}[Z^{b_0} Z^u] \\ &= \lambda_{b_m}^M \lambda_{P_m}^{G_m} \dots \lambda_{P_1}^{G_1} \lambda_{b_0}^S \\ &= \exp\left(-\left(x_{b_0}^S + x_{P_1}^{G_1} + \dots + x_{P_m}^{G_m} + x_{b_m}^M\right)\right) \\ &\equiv \exp(-f(\mathbf{x})). \end{aligned} \quad (\text{A12})$$

The third line is obtained by acting $\tilde{\mathcal{C}}^\dagger$ on Z^{b_m} inside the trace, and then evolve Z^{b_m} backwards according to the definition of $\tilde{\mathcal{C}}$ from Eq. (A10). As a result, we have

$$f(\mathbf{x}) = -\log\left(\sum_{k \in \{0,1\}^n} (-1)^{k \cdot b_m} F_k(\mathbf{x})\right). \quad (\text{A13})$$

This completes the proof of Theorem 6 (b). Recall that this means any rooted cycle f can be learned from a single experiment F . ■

APPENDIX B: PROOF OF THEOREM 8

Proof. We first show that $T_R = \text{Ker}(\mathcal{P} \circ \mathcal{Q})$. For any experiment F , Theorem 6(b) implies that there exists some function $\hat{F}$ such that $F(\mathcal{Q}(\mathbf{r})) = \hat{F}(\mathcal{P} \circ \mathcal{Q}(\mathbf{r}))$ for all

$\mathbf{r} \in X_R$. Thus, for any $\mathbf{t} \in \text{Ker}(\mathcal{P} \circ \mathcal{Q})$,

$$\begin{aligned} F(\mathcal{Q}(\mathbf{r} + \eta\mathbf{t})) &= \hat{F}(\mathcal{P} \circ \mathcal{Q}(\mathbf{r} + \eta\mathbf{t})) \\ &= \hat{F}(\mathcal{P} \circ \mathcal{Q}(\mathbf{r})) = F(\mathcal{Q}(\mathbf{r})), \quad \forall \mathbf{r} \in X_R, \\ &\quad \forall \eta \in \mathbb{R}. \end{aligned} \quad (\text{B1})$$

The second equality uses linearity of $\mathcal{P} \circ \mathcal{Q}$. The above means $\mathbf{t} \in T_R$ by definition. Therefore, $\text{Ker}(\mathcal{P} \circ \mathcal{Q}) \subseteq T_R$; On the other hand, for any $\mathbf{t} \notin \text{Ker}(\mathcal{P} \circ \mathcal{Q})$, $\mathcal{Q}(\mathbf{t}) \notin \text{Ker}(\mathcal{P})$, which means there is at least one f_i such that $f_i(\mathcal{Q}(\mathbf{t})) \neq 0$. Now, Theorem 6(b) says that there exists an experiment F_i and a function $\hat{f}_i$ such that $f_i(\mathcal{Q}(\mathbf{t})) = \hat{f}_i(F_i(\mathcal{Q}(\mathbf{t})))$ for all $\mathbf{r}$. Thus

$$\begin{aligned} f_i(\mathcal{Q}(\mathbf{t})) \neq 0 &\Rightarrow f_i(\mathcal{Q}(\mathbf{r})) \neq f_i(\mathcal{Q}(\mathbf{r} + \mathbf{t})), \quad \forall \mathbf{r} \in X_R, \\ &\Rightarrow \hat{f}_i(F(\mathcal{Q}(\mathbf{r}))) \neq \hat{f}_i(F(\mathcal{Q}(\mathbf{r} + \mathbf{t}))), \quad \forall \mathbf{r} \in X_R, \\ &\Rightarrow F(\mathcal{Q}(\mathbf{r})) \neq F(\mathcal{Q}(\mathbf{r} + \mathbf{t})), \quad \forall \mathbf{r} \in X_R, \end{aligned} \quad (\text{B2})$$

where the first line uses linearity of f_i and $\mathcal{Q}$. This means $\mathbf{t} \notin T_R$. Therefore, $T_R \subseteq \text{Ker}(\mathcal{P} \circ \mathcal{Q})$. This proves that $T_R = \text{Ker}(\mathcal{P} \circ \mathcal{Q})$.

We then show that $L_R = \text{Im}(\mathcal{Q}^\top \circ \mathcal{P}^\top)$. For any $f \in \text{Im}(\mathcal{Q}^\top \circ \mathcal{P}^\top)$, there exists some linear function $h: \mathbb{R}^{|Z|} \mapsto \mathbb{R}$ such that $f(\mathbf{r}) = \mathcal{Q}^\top \circ \mathcal{P}^\top(h)(\mathbf{r}) \equiv h(\mathcal{P} \circ \mathcal{Q}(\mathbf{r}))$, $\forall \mathbf{r} \in X_R$. Theorem 6 implies that each entry of $\mathcal{P}(\mathcal{Q}(\mathbf{r}))$ can be learned from some experiments, thus $f \in L_R$. This means $\text{Im}(\mathcal{Q}^\top \circ \mathcal{P}^\top) \subseteq L_R$; On the other hand, $L_R \perp T_R$, according to similar arguments as in Lemma 3. We have shown $T_R = \text{Ker}(\mathcal{P} \circ \mathcal{Q})$, thus $L_R \subseteq \text{Im}(\mathcal{Q}^\top \circ \mathcal{P}^\top)$ thanks to the orthogonality between kernel and adjoint image. This proves that $L_R = \text{Im}(\mathcal{Q}^\top \circ \mathcal{P}^\top)$. ■

APPENDIX C: PROOF OF THE CONVERSE PART OF THEOREM 10 (b)

Proof. Let $(X_R, \mathcal{Q})$ satisfy a fully local noise model for gate $\mathcal{G}$. Define

$$\begin{aligned} T_0 &:= \{\mathbf{z} \in T : \mathbf{z}^{\mathcal{G}} \in \text{Im} \mathcal{Q}^{\mathcal{G}}\}, \\ D_0 &:= \text{span}\{\mathbf{d}_s : s \subseteq \text{supp}(\mathcal{G}) \text{ or} \\ &\quad s \supseteq \text{supp}(\mathcal{G}) \text{ or } s \cap \text{supp}(\mathcal{G}) = \emptyset\}. \end{aligned} \quad (\text{C1})$$

It has been proved that $T_0 \supseteq D_0$. Now, assume that $\hat{\mathcal{G}}$ induces a connected pattern transfer subgraph. Let $k = |\mathcal{G}|$ and assume $\mathcal{G}$ supports on the first k qubits, without loss of generality. By definition, any $\mathbf{z}^{\mathcal{G}} \in \text{Im} \mathcal{Q}^{\mathcal{G}}$ satisfies

$$z_{b \oplus c_1}^{\mathcal{G}} = z_{b \oplus c_2}^{\mathcal{G}}, \quad \forall b \in \mathbf{P}^k \setminus \{I_k\}, \quad \forall c_1, c_2 \in \mathbf{P}^{n-k}. \quad (\text{C2})$$

Here $\oplus$ means concatenation of bit string, i.e., $P_{b \oplus c} = P_b \otimes P_c$. Equivalently, $\mathbf{z} \cdot (\mathbf{e}_{b \oplus c_1}^{\mathcal{G}} - \mathbf{e}_{b \oplus c_2}^{\mathcal{G}}) = 0$. Now, let us

pick a set of k -qubit Paulis $\{b_j \in \mathbf{P}^k \setminus \{I\}\}_{j=1}^{2^k-2}$ such that the set of edges $\{e_{b_j \oplus 0}^G\}_{j=1}^{2^k-2}$, viewed as undirected, form a spanning tree for the subgraph of vertices $\{t \oplus 0_{n-k} \in \mathbf{V} : t \in \{0, 1\}^k, t \neq 0_k\}$. Such a choice always exists, as we have assumed that the pattern transfer subgraph of $\widehat{G}$ is strongly connected. Next, consider the following vectors:

$$\begin{aligned} \mathbf{k}_{j,c} &:= e_{b_j \oplus I}^G - e_{b_j \oplus c}^G, \quad \forall j = 1, \dots, 2^k - 2, \\ \forall c \in \mathbf{P}^{n-k} \setminus \{I\}. \end{aligned} \quad (\text{C3})$$

It is not hard to see $\{\mathbf{k}_{j,c}\}$ forms a linearly-independent set, as each $e_{b_j \oplus c}^G (c \neq I)$ appears in exactly one $\mathbf{k}_{j,c}$. Let $K := \text{span}\{\mathbf{k}_{j,c}\}$. We thus have

$$\dim(K) = (2^k - 2) \times (2^{n-k} - 1) = 2^n - 2^k - 2^{n-k+1} + 2. \quad (\text{C4})$$

Since K supports on a union of spanning trees, the intersection between K and the cycle space Z must be trivial, i.e., $K \cap Z = \{\mathbf{0}\}$ (see, e.g., Ref. [85]). Meanwhile, one has $T_0 \perp K$ by Eq. (C2) and $T_0 \perp Z$ by the orthogonality between cycle space and cut space. Combining all of these, we have $X \supseteq T_0 \oplus K \oplus Z$, and that

$$\begin{aligned} \dim(T_0) &\leq \dim(X) - \dim(Z) - \dim(K) \\ &= \dim(T) - \dim(K) \\ &= 2^k + 2^{n-k+1} - 3. \end{aligned} \quad (\text{C5})$$

The second line uses the fact that T is the complement of Z in X . On the other hand, by directly counting the number of SDG's in D_0 , we have $\dim(D_0) = (2^k - 2) + 2^{n-k} + (2^{n-k} - 1) \geq \dim(T_0)$. Since we have shown $T_0 \supseteq D_0$, we conclude $D_0 = T_0$. This completes our proof. ■

APPENDIX D: RELATIVE-PRECISION LEARNING FOR TWO-QUBIT GATES IN FULLY LOCAL MODEL

We consider the second type of algorithms that learn gate noise parameters to relative precision. The key is to construct a reduced cycle basis for the gate noise parameter space $L_R^G = \mathcal{Q}^\top(L \cap X^G)$. Note that, one might think

$$\widetilde{L}_R^G := L_R \cap X_R^G = \mathcal{Q}^\top(L) \cap \mathcal{Q}^\top(X^G) \quad (\text{D1})$$

should be the natural definition of the reduced learnable gate parameter space. However, L_R^G and $\widetilde{L}_R^G$ might not be equal in general. Theorem 9 implies the former is the space of functions that can be amplified and learned to relative precision. Nevertheless, in some cases L_R^G and $\widetilde{L}_R^G$ can be shown to be equal.

We focus on the case that every $G \in \mathfrak{G}$ is supported non-trivially on at most two qubits. Since many existing

ALGORITHM 3. Finding a basis for $\widetilde{L}_R^G$.

- (1) Set $A = \{\vartheta_a^G : a \in \mathbf{P}^2 \setminus \{I_2\}, G \in \mathfrak{G}\}$ and $C = \emptyset$.
- (2) Find a non-zero vector $c \in \widetilde{L}_R^G \cap \text{span}(A)$.
- (3) Pick a ϑ with $\vartheta \cdot c \neq 0$, set $A = A \setminus \{\vartheta\}$ and $C = C \cup \{c\}$.
- (4) Repeat steps 2 and 3 until no c can be found, return C .

quantum computing platform uses two-qubit entangling gates such as CZ as a primitive, we expect our method to be practically useful.

To begin with, we will first describe an algorithm for finding a basis for $\widetilde{L}_R^G$ and then show how to augment it to an algorithm for finding reduced cycle basis for $\widetilde{L}_R^G$. Since reduced cycle basis belongs to L_R^G , this implies that $\widetilde{L}_R^G \subseteq L_R^G$. Moreover, by definition $L_R^G \subseteq \widetilde{L}_R^G$, so $\widetilde{L}_R^G = L_R^G$ in this case, and the reduced cycle basis we find for $\widetilde{L}_R^G$ is also a reduced cycle basis for L_R^G .

Lemma D1. Algorithm 3 outputs a basis for $\widetilde{L}_R^G$.

This is a standard linear algebraic procedure. For completeness, a proof is given below.

Proof. We prove the following loop invariant: $(\widetilde{L}_R^G \cap \text{span}(A)) + \text{span}(C) = \widetilde{L}_R^G$.

Initially, $\text{span}(A) = X_R^G \supseteq \widetilde{L}_R^G$ and $\text{span}(C) = \{\mathbf{0}\}$, so $(\widetilde{L}_R^G \cap \text{span}(A)) + \text{span}(C) = \widetilde{L}_R^G$. Now suppose the loop invariant is true before the update step 3. Then $\forall I \in \widetilde{L}_R^G, \exists I_0 \in \widetilde{L}_R^G \cap \text{span}(A)$ and $c_0 \in \text{span}(C)$ s.t. $I = I_0 + c_0$. Then $I_0 - \frac{\vartheta \cdot I_0}{\vartheta \cdot c} c \in \widetilde{L}_R^G \cap \text{span}(A)$. Moreover, $\vartheta \cdot (I_0 - \frac{\vartheta \cdot I_0}{\vartheta \cdot c} c) = 0$, so $I_0 - \frac{\vartheta \cdot I_0}{\vartheta \cdot c} c \in \widetilde{L}_R^G \cap \text{span}(A \setminus \{\vartheta\})$. Since $((\vartheta \cdot I_0)/(\vartheta \cdot c))c + c_0 \in \text{span}(C \cup \{c\})$, by $I = (I_0 - \frac{\vartheta \cdot I_0}{\vartheta \cdot c} c) + (\frac{\vartheta \cdot I_0}{\vartheta \cdot c} c + c_0)$ we have $(\widetilde{L}_R^G \cap \text{span}(A \setminus \{\vartheta\})) + \text{span}(C \cup \{c\}) = \widetilde{L}_R^G$, proving the loop invariant. In the end, $\widetilde{L}_R^G \cap \text{span}(A) = \{\mathbf{0}\}$, so the span of the output C is $\widetilde{L}_R^G$.

Furthermore, for any update step 3, the vector c appended to C has support on ϑ , whereas all the vectors appended afterward do not. Hence C is linearly independent and the output is a basis for $\widetilde{L}_R^G$. ■

Before we augment the algorithm, we first give a characterization of the space $\widetilde{L}_R^G$.

Definition D1. The boundary operator is a linear operator $\partial : X_R^G \rightarrow X_R^S$ defined as

$$\partial(\vartheta_a^G) = \sum_{j \in \text{supp}(G)} \left(\mathbb{1}[\text{pt}(a)_j \neq \mathbf{0}] - \mathbb{1}[\text{pt}(\widehat{G}(a))_j \neq \mathbf{0}] \right) \vartheta_j^S.$$

For $\mathbf{r} \in X_R^G$, $\partial(\mathbf{r})$ is called its boundary.

Lemma D2. $\forall \mathbf{r} \in X_R^G, \mathbf{r} \in \widetilde{L}_R^G$ if and only if $\partial(\mathbf{r}) = \mathbf{0}$.

Proof. Note that any $\mathbf{r} \in X_R^G$ can be expanded as $\mathbf{r} = \sum_{a \in \mathbb{P}^2 \setminus \{I_2\}, \mathcal{G} \in \mathfrak{G}} r_a^G \boldsymbol{\vartheta}_a^G$. We have

$$\begin{aligned}
& \sum_{a \in \mathbb{P}^2 \setminus \{I_2\}, \mathcal{G} \in \mathfrak{G}} r_a^G \boldsymbol{\vartheta}_a^G \in L_R \\
& \Leftrightarrow \sum_{a \in \mathbb{P}^2 \setminus \{I_2\}, \mathcal{G} \in \mathfrak{G}} \sum_{j \in \text{supp}(\mathcal{G})} r_a^G \left(\mathbb{1}[\text{pt}(a)_j \neq \mathbf{0}] \boldsymbol{\vartheta}_j^S + \mathbb{1}[\text{pt}(\hat{\mathcal{G}}(a))_j \neq \mathbf{0}] \boldsymbol{\vartheta}_j^M \right) \in L_R \\
& \Leftrightarrow \sum_{a \in \mathbb{P}^2 \setminus \{I_2\}, \mathcal{G} \in \mathfrak{G}} \sum_{j \in \text{supp}(\mathcal{G})} r_a^G \left(\mathbb{1}[\text{pt}(a)_j \neq \mathbf{0}] - \mathbb{1}[\text{pt}(\hat{\mathcal{G}}(a))_j \neq \mathbf{0}] \right) \boldsymbol{\vartheta}_j^S \in L_R \\
& \Leftrightarrow \sum_{a \in \mathbb{P}^2 \setminus \{I_2\}, \mathcal{G} \in \mathfrak{G}} \sum_{j \in \text{supp}(\mathcal{G})} r_a^G \left(\mathbb{1}[\text{pt}(a)_j \neq \mathbf{0}] - \mathbb{1}[\text{pt}(\hat{\mathcal{G}}(a))_j \neq \mathbf{0}] \right) \boldsymbol{\vartheta}_j^S = \mathbf{0} \\
& \Leftrightarrow \partial(\mathbf{r}) = \mathbf{0}.
\end{aligned} \tag{D2}$$

The second and third lines use the fact that $\mathbf{B}_R$ is a basis for L_R . In the fourth line we use the fact that any non-trivial functions of the reduced state preparation parameters change under some single-qubit depolarization gauge, and is thus not learnable (More formally, $\forall \mathbf{0} \neq \mathbf{f} \in X_R^S, \exists |s| = 1$, s.t., $\mathbf{f} \not\in \mathfrak{D}_{s.}$) ■

Now we augment the algorithm so that it finds a reduced cycle basis. That is, we need to guarantee that every $\mathbf{c}$ appended is a reduced cycle. To do this, we specify how to find $\mathbf{c}$ in step 2 and how to pick $\boldsymbol{\vartheta}$ in step 3. The augmented algorithm consists of 2 stages, for which we will describe how to make the choices separately.

In the first stage, we look for $\mathbf{c} = \sum_{a \in \mathbb{P}^2 \setminus \{I_2\}, \mathcal{G} \in \mathfrak{G}} r_a^G \boldsymbol{\vartheta}_a^G$ with the property that all non-zero r_a^G share the same $\text{supp}(\mathcal{G})$. That is, we group the base vectors $\boldsymbol{\vartheta}_a^G$ by $\text{supp}(\mathcal{G})$ and consider them separately. Fix a common support s , the task simplifies substantially since Lemma D2 simplifies into a requirement on two dimensions (the common support, which is of size 2). Then we may perform the analysis on the pattern transfer subgraph [the union of the pattern transfer subgraphs of $\hat{\mathcal{G}}$ for all $\mathcal{G} \in \mathfrak{G}$ with $\text{supp}(\mathcal{G}) = s$]. In the following we consider the possible shapes of the pattern transfer subgraph. Since self loops can be identified and picked as reduced cycle basis easily at this stage, we do not need to explicitly consider them.

Lemma D3. For two qubit Clifford gates, there are three types of non-trivial pattern transfer subgraphs, the CNOT-like, SWAP-like, and *i*SWAP-like, as shown in Fig. 11.

Proof. By Theorem 26 of Ref. [86], any two qubits Clifford gates is equivalent to at most three layers of gates (each layer may or may not exist): (1) a tensor of single-qubit Clifford gates $G \otimes H$; (2) a SWAP gate; (3) a generalized CNOT gate $(I \otimes I + P \otimes I + I \otimes Q - P \otimes Q)/2$. Since single-qubit Clifford gates before or after a two-qubit

Clifford gate does not change the pattern transfer graph, in the pattern transfer graph sense, any two-qubit Clifford gates is equivalent to at most two layers of gates: (1) a SWAP gate; (2) a CNOT gate. When both gates are absent, the pattern transfer graph is trivial (in this case the Clifford gate is a tensor product of single qubit Clifford gates, which by assumption should not be considered as noisy). If only CNOT is present, the pattern transfer graph is CNOT-like. If only SWAP is present, the pattern transfer graph is SWAP-like. If both gates are present, the pattern transfer graph is *i*SWAP-like. ■

From Fig. 11 one can easily see that the union of pattern transfer subgraphs still has only three possible shapes. The important message of Lemma D3 is that for any edge in the pattern transfer subgraph, there is another edge of the reverse direction.

Now we resume the description of the first stage. Obviously the cycles in the pattern transfer subgraphs all corresponds to reduced cycles (by considering Pauli operators with the same support). One can easily find a cycle basis for the cycle space of the pattern transfer subgraph. Additionally, the cycle basis can be chosen such that for each edge $\boldsymbol{\vartheta}_i$ in the pattern transfer subgraph, there exists a length 2 reduced cycle that uses it. That is, we have picked a reduced cycle $\boldsymbol{\vartheta}_i + \boldsymbol{\vartheta}'_i$ in the basis, where $\boldsymbol{\vartheta}'_i$ is an edge with the reverse direction of $\boldsymbol{\vartheta}_i$. In the first stage of the algorithm, we just pick $\mathbf{c}$ to be such a cycle basis. They are naturally reduced cycles. This can always be done due to the vast freedom in the choice of $\boldsymbol{\vartheta}$. For the pattern transfer

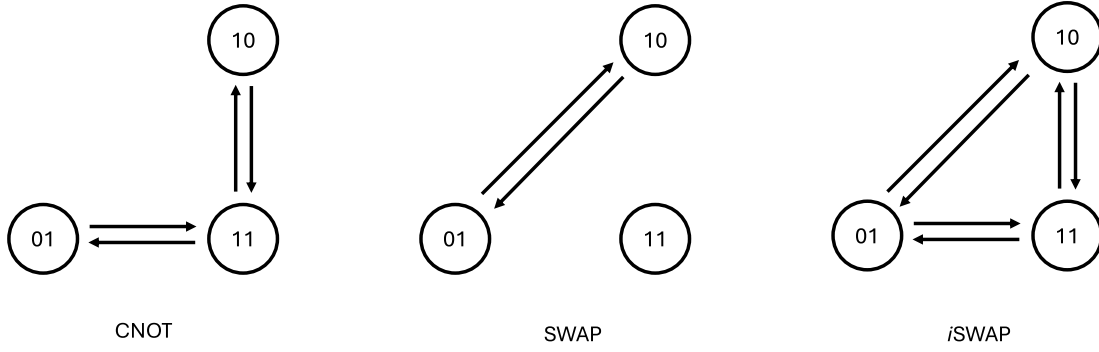


FIG. 11. The possible pattern transfer subgraphs of non-trivial two-qubit Clifford gates. Self loops are omitted. Multi-edges are shown by single edges.

subgraph, at each iteration we eliminate an edge from A . In the end the remaining edges form a spanning tree. Again due to the freedom in the choice of $\boldsymbol{\vartheta}$ we are free to choose which spanning tree should be kept in A and direction of the edges of the remaining spanning tree. For CNOT-like and SWAP-like pattern transfer subgraphs, there is only one possible shape of the spanning tree. As for iSWAP-like pattern transfer graph, wlog we pick (01, 11) and (10, 11) as the spanning tree. The edge orientation for the spanning trees could be arbitrary.

Now the algorithm proceeds to the second stage, where we find reduced cycles across different supports. This stage is illustrated in Fig. 12. According to Lemma D2 we associate the base vectors in A with their boundaries. Thanks to Lemma D3, there are only two types of boundaries: Type 1 boundary takes ± 1 at one entry and 0 at other places, corresponding to edges (10, 11) or (01, 11) in Fig. 11. Type 2 boundary takes 1 at one entry, -1 at another entry, and 0 at other places, corresponding to edge (10, 01) in Fig. 11. Moreover, if we consider the edge of reverse direction, the corresponding boundary is negated. In the following we will find some candidate $\boldsymbol{c} = \sum \text{sgn}_i \boldsymbol{\vartheta}_i$ and then prove that it is a reduced cycle when all the signs are positive. When some signs are negative, we revert some edges to make the signs positive. This is actually a simplification and hides the following fact: though $\boldsymbol{c}$ itself is not a reduced cycle, $\tilde{\boldsymbol{c}} = \boldsymbol{c} + \sum_{\text{sgn}_i = -1} (\boldsymbol{\vartheta}_i + \boldsymbol{\vartheta}'_i)$ ($\boldsymbol{c}$ with edge directions reverted) is a reduced cycle. Moreover, since the $\boldsymbol{\vartheta}_i + \boldsymbol{\vartheta}'_i$ s are already in C , putting $\boldsymbol{c}$ or $\tilde{\boldsymbol{c}}$ into C is equivalent in terms of making C a basis for $\widetilde{L}_R^G$.

According to the proof of Lemma D1, if there are remaining degree of freedom in $\widetilde{L}_R^G$, there should exist a non-zero vector $\boldsymbol{c}' \in \widetilde{L}_R^G \cap \text{supp}(A)$. Then $\partial(\boldsymbol{c}') = \mathbf{0}$. We can find such $\boldsymbol{c}' = \sum_i r_{a_i}^{\mathcal{G}_i} \boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i}$ (the summation is only over non-zero terms) via simple linear algebra on the boundaries of the edges in A . The $r_{a_i}^{\mathcal{G}_i}$ found is rational and thus can be made integral by multiplying them by the least common multiple of the denominators.

We now explain how to find a reduced cycle $\boldsymbol{c} \in L_R^G$ based on $\boldsymbol{c}'$. Expand $\partial(\boldsymbol{c}') = \sum_i r_{a_i}^{\mathcal{G}_i} \partial(\boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i}) = \mathbf{0}$. Among the summation, if there are two type-1 boundaries with non-zero entry at the same qubit [reverting the edge directions if needed, wlog assume their sum is $\mathbf{0}$: $\partial(\boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i}) + \partial(\boldsymbol{\vartheta}_{a_j}^{\mathcal{G}_j}) = \mathbf{0}$], then $\boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i} + \boldsymbol{\vartheta}_{a_j}^{\mathcal{G}_j}$ is a reduced cycle that can be chosen as $\boldsymbol{c}$. To see this, first observe that $\text{supp}(\mathcal{G}_i) \neq \text{supp}(\mathcal{G}_j)$, since otherwise $\boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i}$ and $\boldsymbol{\vartheta}_{a_j}^{\mathcal{G}_j}$ are two edges of a spanning tree. But $\text{supp}(\mathcal{G}_i)$ and $\text{supp}(\mathcal{G}_j)$ intersect at the shared non-zero entry of the two boundaries. So wlog assume $\text{supp}(\mathcal{G}_i) = \{1, 2\}$, $\text{supp}(\mathcal{G}_j) = \{2, 3\}$, $\partial(\boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i}) = [0, 1, 0, \dots]$ and $\partial(\boldsymbol{\vartheta}_{a_j}^{\mathcal{G}_j}) = [0, -1, 0, \dots]$. Then $\boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i} + \boldsymbol{\vartheta}_{a_j}^{\mathcal{G}_j} = \mathcal{Q}^T(\boldsymbol{e}_{(a_i)_{12}X_3I_4 \dots I_n}^{\mathcal{G}_i} + \boldsymbol{e}_{X_1(a_j)_{23}I_4 \dots I_n}^{\mathcal{G}_j})$, so it is a reduced cycle.

If there are no two type-1 boundaries with the same non-zero entry, since the sum of the boundaries is $\mathbf{0}$, there must be some type-2 boundaries in the summation. Pick a type-2 boundary $r_{a_0}^{\mathcal{G}_0} \boldsymbol{\vartheta}_{a_0}^{\mathcal{G}_0}$. This vector has a positive entry. So there must be another boundary $r_{a_1}^{\mathcal{G}_1} \boldsymbol{\vartheta}_{a_1}^{\mathcal{G}_1}$ that is negative on the entry. Either it is a type-1 boundary and the procedure terminates here, or it is a type-2 boundary and it has another positive entry, and we repeat this chasing procedure. The procedure cannot last infinitely, so at some point we must terminate at a type-1 boundary, or we have chased back to an entry we have already considered. We do the same chasing procedure for the negative entry of $r_{a_0}^{\mathcal{G}_0} \boldsymbol{\vartheta}_{a_0}^{\mathcal{G}_0}$.

If both chasing procedure terminates with a type-1 boundary, we get the following vector of boundary $\mathbf{0}$: $\sum_{i=-t_1}^{t_2} \text{sgn}(r_{a_i}^{\mathcal{G}_i}) \boldsymbol{\vartheta}_{a_i}^{\mathcal{G}_i}$, where $\boldsymbol{\vartheta}_{a_{-t_1}}^{\mathcal{G}_{-t_1}}$ and $\boldsymbol{\vartheta}_{a_{t_2}}^{\mathcal{G}_{t_2}}$ have type-1 boundaries and the others have type-2 boundaries. Reverting the directions of basis if necessary, we can further assume all the signs are $+1$. Then we get a reduced cycle that can be used as $\boldsymbol{c}$. To see this, note that since type-2 boundaries only come from SWAP-like pattern transfer graph, different type-2 boundaries have different supports (type-2 boundaries have the same support as the corresponding $\mathcal{G}$). Let $m = t_1 + t_2 - 1$.

(a)

$$\begin{aligned} 0 &= r_{a_1}^{\mathcal{G}_1} \partial(\vartheta_{a_1}^{\mathcal{G}_1}) + r_{a_2}^{\mathcal{G}_2} \partial(\vartheta_{a_2}^{\mathcal{G}_2}) + r_{a_3}^{\mathcal{G}_3} \partial(\vartheta_{a_3}^{\mathcal{G}_3}) + r_{a_4}^{\mathcal{G}_4} \partial(\vartheta_{a_4}^{\mathcal{G}_4}) \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \\ &\Downarrow \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

There exists two type-1 boundaries with the same support.

(b)

$$\begin{aligned} 0 &= r_{a_{-1}}^{\mathcal{G}_{-1}} \partial(\vartheta_{a_{-1}}^{\mathcal{G}_{-1}}) + r_{a_0}^{\mathcal{G}_0} \partial(\vartheta_{a_0}^{\mathcal{G}_0}) + r_{a_1}^{\mathcal{G}_1} \partial(\vartheta_{a_1}^{\mathcal{G}_1}) + r_{a_2}^{\mathcal{G}_2} \partial(\vartheta_{a_2}^{\mathcal{G}_2}) \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \\ &\Downarrow \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \end{aligned}$$

Chasing procedure at both end terminates with a type-1 boundary.

(c)

$$\begin{aligned} 0 &= r_{a_{-1}}^{\mathcal{G}_{-1}} \partial(\vartheta_{a_{-1}}^{\mathcal{G}_{-1}}) + r_{a_0}^{\mathcal{G}_0} \partial(\vartheta_{a_0}^{\mathcal{G}_0}) + r_{a_1}^{\mathcal{G}_1} \partial(\vartheta_{a_1}^{\mathcal{G}_1}) + r_{a_2}^{\mathcal{G}_2} \partial(\vartheta_{a_2}^{\mathcal{G}_2}) + r_{a_3}^{\mathcal{G}_3} \partial(\vartheta_{a_3}^{\mathcal{G}_3}) + r_{a_4}^{\mathcal{G}_4} \partial(\vartheta_{a_4}^{\mathcal{G}_4}) \\ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} \\ &\Downarrow \\ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

A chasing procedure ends with a loop.

FIG. 12. Illustration of the second stage of the algorithm. In each subgraph, the upper equation stands for some c' such that $\partial(c') = 0$ while the bottom stands for a reduced cycle $c \in L_R^G$ constructed based on c' . Concrete examples of c corresponding to each class is given in Fig. 13. (a) There exists two type-1 boundaries with the same support. (b) Chasing procedure at both end terminates with a type-1 boundary. (c) A chasing procedure ends with a loop.

Wlog we assume the support of $\mathcal{G}_{-t_1+1}, \dots, \mathcal{G}_{t_2-1}$ are $\{1, 2\}, \{2, 3\}, \{3, 4\}, \dots, \{m, m+1\}$, and that $\partial(\vartheta_{a_{-t_1}}^{\mathcal{G}_{-t_1}}) = [1, 0, \dots, 0]$, $\partial(\vartheta_{a_{t_2}}^{\mathcal{G}_{t_2}}) = [\dots, 0, -1, 0, \dots]$. Then,

$$\begin{aligned} \sum_{i=-t_1}^{t_2} \vartheta_{a_i}^{\mathcal{G}_i} &= \mathcal{Q}^T \left(e_{P_1 Q_2 Q_3 \dots Q_n}^{\mathcal{G}_{-t_1}} + e_{(a_{-t_1}+1)_{12} Q_3 \dots Q_n}^{\mathcal{G}_{-t_1+1}} + e_{Q_1(a_{-t_1}+2)_{23} Q_3 \dots Q_n}^{\mathcal{G}_{-t_1+2}} \right. \\ &\quad \left. + \dots + e_{Q_1 \dots Q_{m-1}(a_{t_2}-1)_{m,m+1} Q_{m+2} \dots Q_n}^{\mathcal{G}_{t_2-1}} + e_{Q'_1 \dots Q'_m I_{m+1} Q'_{m+2} \dots Q'_n}^{\mathcal{G}_{t_2}} \right), \end{aligned} \quad (\text{D3})$$

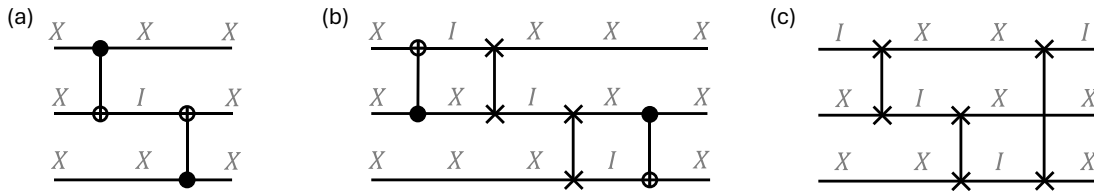


FIG. 13. Concrete examples of three possible classes of reduced cycles involving more than one gate, corresponding to Fig. 12. The two-qubit gates appearing here are CNOT and SWAP.

where we use $a_{-t_1} = P_1 Q_k$ and $a_{t_2} = Q_{k'} I_{m+1}$, for some indexes k, k' and non-trivial single-qubit Pauli operators P, Q, Q' . Note that the patterns of $a_{-t_1+1}, \dots, a_{t_2-1}$ are all 01, according to their corresponding boundaries. It is not hard to see the above equation represents the action of a sequence of SWAP-like gates, thus yields a reduced cycle which can be chosen as c . An example is given in Fig. 12(b).

Finally, consider the case where some chasing procedure ends with a loop. That is, we have found a cycle of type-2 boundaries $\sum_{i=1}^m \text{sgn}(r_{a_i}^{\mathcal{G}_i}) \mathfrak{P}_{a_i}^{\mathcal{G}_i}$. Revert directions if necessary, wlog we can assume that all the signs are positive. Wlog assume the support of $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_m$ are $\{1, 2\}, \{2, 3\}, \dots, \{m, 1\}$, and that $\partial(\mathfrak{P}_{a_1}^{\mathcal{G}_1}) = [-1, 1, 0, \dots]$. Then,

$$\sum_i \mathfrak{P}_{a_i}^{\mathcal{G}_i} = \mathcal{Q}^T \left(e_{(a_1)_{12} X_3 X_4 \dots X_n}^{\mathcal{G}_1} + e_{X_1 (a_2)_{23} X_4 \dots X_n}^{\mathcal{G}_2} + \dots + e_{X_2 \dots X_{m-1} (a_m)_{m,1} X_{m+1} \dots X_n}^{\mathcal{G}_m} \right) \quad (\text{D4})$$

so it is a reduced cycle. Note that the patterns of $a_1, \dots, a_m$ are all 01 according to their corresponding boundaries. This case is illustrated in Fig. 12(c). Thus we see that as long as there are remaining degree of freedom in $\widetilde{L}_R^G$, we can find a reduced cycle, so Algorithm 3 can terminate correctly. Since all the basis vectors are also elements of L_R^G , we have proved in this case $L_R^G = \widetilde{L}_R^G$.

APPENDIX E: ABOUT QUASI-LOCAL PAULI CHANNELS

Throughout this section, when we talk about a Pauli channel, we always assume all of its eigenvalues to be strictly positive, so that $x_a = -\log \lambda_a$ is well-defined.

Lemma E1. Let $\Omega = 2^{[n]} \setminus \{\emptyset\}$. Then any n -qubit Pauli channel is Ω -local.

Proof. For any set of parameters $\{x_a : a \in \mathbf{P}^n\}$, we claim there always exist $\{r_b : b \in \mathbf{P}^n\}$ such that,

$$x_a = \sum_{b \triangleleft a} r_b, \quad \forall a \in \mathbf{P}^n. \quad (\text{E1})$$

$$r_b = \sum_{a \triangleleft b} (-1)^{w(a)-w(b)} x_a, \quad \forall b \in \mathbf{P}^n. \quad (\text{E2})$$

Indeed, substituting the second expression into RHS of the first yields

$$\begin{aligned} \sum_{b \triangleleft a} \sum_{c \triangleleft b} (-1)^{w(c)-w(b)} x_c &= \sum_{c \triangleleft a} x_c \sum_{b: c \triangleleft b \triangleleft a} (-1)^{w(c)-w(b)} \\ &= \sum_{c \triangleleft a} x_c \sum_{k=0}^{|w(a)-w(c)|} \binom{w(a)-w(c)}{k} \\ &\quad \times (-1)^k \\ &= \sum_{c \triangleleft a} x_c \mathbb{1}[w(a) = w(c)] \\ &= x_a. \end{aligned} \quad (\text{E3})$$

The first line changes the order of summation. The third line uses the binomial theorem. Further note that $x_{I_n} = 0 \Leftrightarrow r_{I_n} = 0$. Since any Pauli channel is trace-preserving (i.e., $x_{I_n} = 0$), it is $2^{[n]} \setminus \{\emptyset\}$ -local by choosing r_b according to Eq. (E2). ■

The following lemma explains the equivalence between the quasi-local Pauli channels we consider and the sparse Pauli-Lindblad noise model. We note that basically the same results are given by Ref. [17, Theorem SIV.1].

Lemma E2. Given an n -qubit Pauli channel Λ and a factor set Ω , if Λ can be written in the following concatenating form:

$$\Lambda(\cdot) = \bigcirc_{b \sim \Omega} (\eta_b P_b(\cdot) P_b + (1 - \eta_b)(\cdot)), \quad (\text{E4})$$

with real parameters $\{\eta_b\}_{b \sim \Omega}$ such that $\eta_b < 1/2, \forall b \neq I_n$, then Λ is Ω -local. Conversely, any Ω -local Pauli channel can be written in the above form.

Proof. Suppose Eq. (E4) holds. By definition of the Pauli eigenvalues,

$$\lambda_a = \text{Tr}[P_a \Lambda(P_a)]/2^n = \prod_{b \sim \Omega} (1 - 2\eta_b)^{\langle a, b \rangle}. \quad (\text{E5})$$

Taking a log on both sides yields (recall that we require $\eta_b < 1/2$)

$$\begin{aligned} x_a &:= -\log(\lambda_a) = -\sum_{b \sim \Omega} \langle a, b \rangle \log(1 - 2\eta_b) \\ &=: \sum_{b \sim \Omega} \langle a, b \rangle \tau_b, \end{aligned} \quad (\text{E6})$$

where we define $\tau_b := -\log(1 - 2\eta_b)$. On the other hand, recall $\{r_b\}$ as defined in Eq. (E2). Substituting Eq. (E6) into

the expression of r_b yields,

$$r_b = \sum_{c \sim \Omega} \sum_{a \triangleleft b} (-1)^{w(a)-w(b)} \langle a, c \rangle \tau_c. \quad (\text{E7})$$

Now we claim that $r_b = 0$ unless $b \sim \Omega$. If the condition does not hold, for any $c \sim \Omega$, there is at least one $i \in [n]$ such that $c_i = I$ but $b_i \neq I$. For any $a \triangleleft b$ such that $a_i = I$, let $\tilde{a}$ be obtained by replacing the i th entry of a by b_i . Then we obviously have

$$(-1)^{w(a)-w(b)} \langle a, c \rangle + (-1)^{w(\tilde{a})-w(b)} \langle \tilde{a}, c \rangle = 0.$$

Further note that $\{a \in \mathbf{P}^n : a \triangleleft b\}$ can be decomposed into disjoint pairs of the above form $(a, \tilde{a})$, we thus conclude that r_b must be 0, justifying our claim. We can thus rewrite the expression for $\mathbf{x}$ as

$$x_a = \sum_{b \triangleleft a, b \sim \Omega} r_b, \quad (\text{E8})$$

depending on a set of real parameters $\{r_b\}_{b \sim \Omega}$. Exponentiating both sides yields,

$$\lambda_a = \prod_{b \triangleleft a, b \sim \Omega} \exp(-r_b). \quad (\text{E9})$$

This completes the proof that Λ is Ω -local.

Conversely, let us assume that Λ is Ω -local, i.e., Eq. (E8) holds. Define τ_b according to

$$\tau_b := \mathbb{1}[b \neq I_n] \cdot \frac{-2}{4^n} \sum_a (-1)^{\langle a, b \rangle} x_a, \quad \forall b \in \mathbf{P}^n, \quad (\text{E10})$$

which gives us $x_a = \sum_b \langle a, b \rangle \tau_b$. Indeed,

$$\begin{aligned} \sum_b \langle a, b \rangle \tau_b &= -\frac{2}{4^n} \sum_{b \neq I_n} \langle a, b \rangle \sum_c (-1)^{\langle c, b \rangle} x_c \\ &= -\frac{2}{4^n} \sum_{b \neq I_n} \frac{1 - (-1)^{\langle a, b \rangle}}{2} \sum_c (-1)^{\langle c, b \rangle} x_c \\ &= \frac{1}{4^n} \sum_c x_c \sum_{b \neq I_n} (-(-1)^{\langle c, b \rangle} + (-1)^{\langle a+c, b \rangle}) \\ &= \sum_c x_c (-\mathbb{1}[c = I_n] + \mathbb{1}[c = a]) \\ &= x_a - x_{I_n}, \end{aligned} \quad (\text{E11})$$

which is just x_a since $x_{I_n} = 0$ by the trace preserving condition. Now substituting Eq. (E8) into Eq. (E10), considering

only $b \neq I_n$,

$$\begin{aligned} \tau_b &= -\frac{2}{4^n} \sum_a (-1)^{\langle a, b \rangle} \sum_{c \triangleleft a, c \sim \Omega} r_c \\ &= -\frac{2}{4^n} \sum_{c \sim \Omega} r_c \sum_{a: c \triangleleft a} (-1)^{\langle a, b \rangle}. \end{aligned} \quad (\text{E12})$$

We claim that $\tau_b = 0$ unless $b \sim \Omega$. Similar to the previous case, if $b \sim \Omega$ does not hold, for any $c \sim \Omega$ there exists an index i such that $c_i = I$ but $b_i \neq I$. Consequently, exactly half of the Pauli operators from $\{a : c \triangleleft a\}$ commute with b while the other half anti-commute with b , depending on whether a_i commutes with b_i . Thus, the above sum always yields zero, proving our claim. By setting $\eta_b = (1 - \exp(-\tau_b))/2$ for all $b \sim \Omega$, we see Eq. (E4) is a valid representation for Λ . This completes our proof. ■

We are now ready to give a proof for Lemma 10.

Proof for Lemma 10. Thanks to Lemma E2, any Ω -local Pauli channel Λ can be written in the form of Eq. (E4). Conjugate Λ with $\mathcal{G}$ yields,

$$\begin{aligned} \mathcal{G} \Lambda \mathcal{G}^\dagger &= \bigcirc_{b \sim \Omega} (\eta_b \mathcal{G} (P_b(\mathcal{G}^\dagger(\cdot)) P_b) + (1 - \eta_b)(\cdot)) \\ &= \bigcirc_{b \sim \Omega} (\eta_b P_{\mathcal{G}(b)}(\cdot) P_{\mathcal{G}(b)} + (1 - \eta_b)(\cdot)). \end{aligned} \quad (\text{E13})$$

If one can show that $\mathcal{G}$ is a permutation acting on $\{b : b \sim \Omega\}$, by simple relabeling one can see $\mathcal{G} \Lambda \mathcal{G}^\dagger$ also has the form of Eq. (E4) and is thus Ω -local. For any $b \sim \Omega$, there must be a $v \in \Omega$ such that $\text{supp}(b) = v$. it is not hard to see that

$$\text{supp}(\mathcal{G}(b)) \subseteq \Xi_{\mathcal{G}}(v). \quad (\text{E14})$$

By assumption, $\Xi_{\mathcal{G}}(v) \in \Omega$, thus $\text{supp}(\mathcal{G}(b)) \in \Omega$ by the property of Ω , which means $\mathcal{G}(b) \sim \Omega$. Since $\mathcal{G}$ is invertible, this completes the proof that $\mathcal{G}$ acts as a permutation over $\{b : b \sim \Omega\}$, and thus proving $\mathcal{G} \Lambda \mathcal{G}^\dagger$ is Ω -local. Since the same argument holds by replacing $\mathcal{G}$ with $\mathcal{G}^\dagger$, this completes the proof of Lemma 10. ■

APPENDIX F: ADDITIONAL PROOFS FOR SEC. VII C

1. Gauge for nearest-neighbor CZ

Proposition F1. For the gate set and noise model described in Sec. VII C, the embedded gauge space is given by $\mathcal{Q}(T_R) = \text{span}\{\mathfrak{d}_{\{i\}} : i \in [n]\}$.

Proof. Since both state preparation and measurement noise are Ω -local, Theorem 11 (a) says,

$$\{\mathbf{z} \in T : \pi^S(\mathbf{z}) \in \text{Im } \mathcal{Q}^S, \pi^M(\mathbf{z}) \in \text{Im } \mathcal{Q}^M\} = \text{span}\{\mathfrak{d}_v : v \in \Omega\} = \text{span}\{\mathfrak{d}_{\{i\}}, \mathfrak{d}_{\{i,i+1\}} : i \in [n]\}. \quad (\text{F1})$$

Then, by Lemma 6,

$$\begin{aligned} \mathcal{Q}(T_R) &= \text{span}\{\mathfrak{d}_{\{i\}}, \mathfrak{d}_{\{i,i+1\}} : i \in [n]\} \cap \{\mathbf{z} \in T : \pi^{\mathcal{G}_e}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_e}, \pi^{\mathcal{G}_o}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_o}\} \\ &= \{\mathbf{z} \in \text{span}\{\mathfrak{d}_{\{i\}}, \mathfrak{d}_{\{i,i+1\}} : i \in [n]\} : \pi^{\mathcal{G}_e}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_e}, \pi^{\mathcal{G}_o}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_o}\}. \end{aligned} \quad (\text{F2})$$

We are going to show that

$$\{\mathbf{z} \in \text{span}\{\mathfrak{d}_{\{i\}}, \mathfrak{d}_{\{i,i+1\}} : i \in [n]\} : \pi^{\mathcal{G}_e}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_e}\} = \text{span}\{\{\mathfrak{d}_{\{i\}} : i \in [n]\} \cup \{\mathfrak{d}_{\{2k-1,2k\}} : k \in [n/2]\}\}, \quad (\text{F3})$$

$$\{\mathbf{z} \in \text{span}\{\mathfrak{d}_{\{i\}}, \mathfrak{d}_{\{i,i+1\}} : i \in [n]\} : \pi^{\mathcal{G}_o}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_o}\} = \text{span}\{\{\mathfrak{d}_{\{i\}} : i \in [n]\} \cup \{\mathfrak{d}_{\{2k,2k+1\}} : k \in [n/2]\}\}, \quad (\text{F4})$$

which would then implies $\mathcal{Q}(T_R) = \text{span}\{\mathfrak{d}_{\{i\}} : i \in [n]\}$. Let us look at Eq. (F3) first. To see $\text{LHS} \supseteq \text{RHS}$, note that $\mathcal{G}_e^\dagger \mathcal{D}_{\{i\}} \mathcal{G}_e \mathcal{D}_{\{i\}}^{-1}$ and $\mathcal{G}_e^\dagger \mathcal{D}_{\{2k-1,2k\}} \mathcal{G}_e \mathcal{D}_{\{2k-1,2k\}}^{-1}$ are both clearly Ω -local for all $i \in [n], k \in [n/2]$. Here $\mathcal{D}_v$ is the subsystem depolarizing channel on v defined in Eq. (40). Using the same argument as in the proof of Theorem 11 (b), we can conclude those SDGs are indeed allowed. To see $\text{LHS} \subseteq \text{RHS}$, it remains to show that $\text{span}\{\mathfrak{d}_{\{2k,2k+1\}} : k \in [n/2]\} \cap \text{LHS} = \{\mathbf{0}\}$. Consider any vector of the form $\mathbf{z} = \sum_{k \in [n/2]} \alpha_k \mathfrak{d}_{\{2k,2k+1\}}$. By definition of the SDG,

$$z_a^{\mathcal{G}_e} := \mathbf{e}_a^{\mathcal{G}_e} \cdot \mathbf{z} = \sum_{k \in [n/2]} \alpha_k (\mathbb{1}[\text{pt}(\mathcal{G}(a))_{\{2k,2k+1\}} \neq 00] - \mathbb{1}[\text{pt}(a)_{\{2k,2k+1\}} \neq 00]). \quad (\text{F5})$$

Suppose $\pi^{\mathcal{G}_e}(\mathbf{z}) \in \text{Im } \mathcal{Q}^{\mathcal{G}_e}$, then $z_a^{\mathcal{G}_e} = \sum_{b \in \Omega} t_b$ for some parameters t_b . Specifically, for any $k \in [n/2]$, if we choose $a = X_{2k-1}X_{2k+3}$, since $\{2k-1, 2k+3\} \notin \Omega$ given that $n \geq 6$, we have

$$z_{X_{2k-1}X_{2k+3}}^{\mathcal{G}_e} = z_{X_{2k-1}}^{\mathcal{G}_e} + z_{X_{2k+3}}^{\mathcal{G}_e}. \quad (\text{F6})$$

On the other hand, Eq. (F5) gives (see Fig. 14 for how these are computed)

$$\begin{aligned} z_{X_{2k-1}X_{2k+3}}^{\mathcal{G}_e} &= \alpha_{k-1}(1-1) + \alpha_k(1-0) + \alpha_{k+1}(1-1) = \alpha_k. \\ z_{X_{2k-1}}^{\mathcal{G}_e} &= \alpha_{k-1}(1-1) + \alpha_k(1-0) = \alpha_k. \\ z_{X_{2k+3}}^{\mathcal{G}_e} &= \alpha_{k+1}(1-1) + \alpha_k(1-0) = \alpha_k. \end{aligned} \quad (\text{F7})$$

But this means α_k must be 0, which finishes the proof that $\text{span}\{\mathfrak{d}_{\{2k,2k+1\}} : k \in [n/2]\} \cap \text{LHS} = \{\mathbf{0}\}$. Equation (F4) can be proved similarly by cyclically shifting every index by 1. Combining everything, we complete the proof that $\mathcal{Q}(T_R) = \text{span}\{\mathfrak{d}_{\{i\}} : i \in [n]\}$. ■

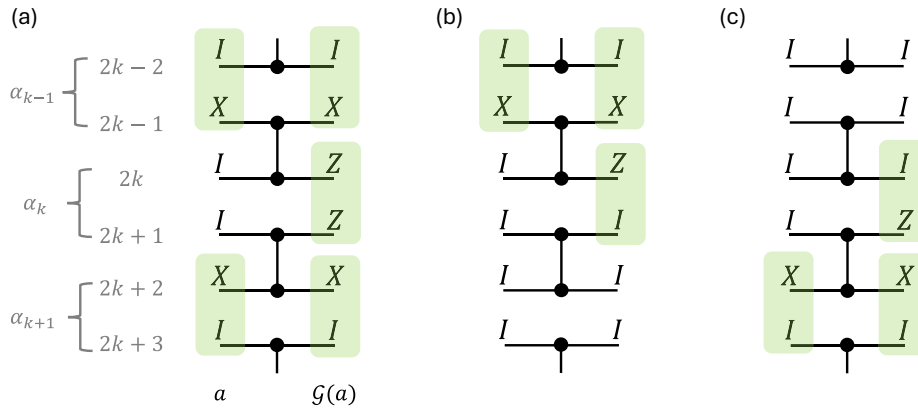


FIG. 14. Illustration for calculations of $z_a^{\mathcal{G}_e}$ where (a) $a = X_{2k-1}X_{2k+3}$, (b) $a = X_{2k-1}$, and (c) $a = X_{2k+3}$, respectively. The labels of qubits and the corresponding coefficient α_k is shown in the leftmost side. The shaded region represents a non-zero pattern in each $\{2k, 2k+1\}$ subsystem, for both a and $\mathcal{G}_e(a)$.

2. Cycle basis for the nearest-neighbor CZ

To prove that Eqs. (84), (85), (86) constitute a basis for L_R^G , it suffices to prove that they form a complete basis for $\widetilde{L}_R^G$ since $L_R^G \subseteq \widetilde{L}_R^G$. We may further perform an invertible linear operation on them first.

Take $\mathcal{Q}^\top \left(e_{I_1 X_2 I_3 X_4 I_5}^{\mathcal{G}_o} + e_{I_1 X_2 Z_3 X_4 Z_5}^{\mathcal{G}_e} + e_{Z_1 X_2 I_3 X_4 Z_5}^{\mathcal{G}_o} + e_{Z_1 X_2 Z_3 X_4 I_5}^{\mathcal{G}_e} \right)$ from Eq. (86) as an example. We change it into

$$\begin{aligned}
& \frac{1}{2} \mathcal{Q}^\top \left(e_{I_1 X_2 I_3 X_4 I_5}^{\mathcal{G}_o} + e_{I_1 X_2 Z_3 X_4 Z_5}^{\mathcal{G}_e} + e_{Z_1 X_2 I_3 X_4 Z_5}^{\mathcal{G}_o} + e_{Z_1 X_2 Z_3 X_4 I_5}^{\mathcal{G}_e} \right) - \frac{1}{4} \mathcal{Q}^\top \left(e_{Z_1 X_2 I_3}^{\mathcal{G}_o} + e_{Z_1 X_2 Z_3}^{\mathcal{G}_o} \right) \\
& - \frac{1}{2} \mathcal{Q}^\top \left(e_{I_1 X_2 Z_3}^{\mathcal{G}_e} + e_{Z_1 X_2 Z_3}^{\mathcal{G}_e} \right) - \frac{1}{4} \mathcal{Q}^\top \left(e_{I_3 X_4 Z_5}^{\mathcal{G}_e} + e_{Z_3 X_4 Z_5}^{\mathcal{G}_e} \right) + \frac{1}{4} \mathcal{Q}^\top \left(e_{X_2 Z_3}^{\mathcal{G}_o} + e_{X_2 I_3}^{\mathcal{G}_o} \right) \\
& - \frac{1}{2} \mathcal{Q}^\top \left(e_{X_4 Z_5}^{\mathcal{G}_o} + e_{X_4 I_5}^{\mathcal{G}_o} \right) - \frac{3}{4} \mathcal{Q}^\top \left(e_{Z_3 X_4}^{\mathcal{G}_e} + e_{I_3 X_4}^{\mathcal{G}_e} \right) + \mathcal{Q}^\top \left(e_{Z_3 I_4}^{\mathcal{G}_e} \right) \\
& = \frac{1}{2} \left(\vartheta_{Z_1}^{\mathcal{G}_o} + 2\vartheta_{X_2}^{\mathcal{G}_o} + 2\vartheta_{X_4}^{\mathcal{G}_o} + \vartheta_{Z_5}^{\mathcal{G}_o} + \vartheta_{Z_1 X_2}^{\mathcal{G}_o} + \vartheta_{X_4 Z_5}^{\mathcal{G}_o} + \vartheta_{Z_1}^{\mathcal{G}_e} + 2\vartheta_{X_2}^{\mathcal{G}_e} + 2\vartheta_{Z_3}^{\mathcal{G}_e} + 2\vartheta_{X_4}^{\mathcal{G}_e} + \vartheta_{Z_5}^{\mathcal{G}_e} \right. \\
& \quad \left. + \vartheta_{Z_1 X_2}^{\mathcal{G}_e} + 2\vartheta_{X_2 Z_3}^{\mathcal{G}_e} + 2\vartheta_{Z_3 X_4}^{\mathcal{G}_e} + \vartheta_{X_4 Z_5}^{\mathcal{G}_e} \right) - \frac{1}{4} \left(2\vartheta_{Z_1}^{\mathcal{G}_o} + 2\vartheta_{X_2}^{\mathcal{G}_o} + \vartheta_{Z_3}^{\mathcal{G}_o} + 2\vartheta_{Z_1 X_2}^{\mathcal{G}_o} + \vartheta_{X_2 Z_3}^{\mathcal{G}_o} \right) \\
& - \frac{1}{2} \left(\vartheta_{Z_1}^{\mathcal{G}_e} + 2\vartheta_{X_2}^{\mathcal{G}_e} + 2\vartheta_{Z_3}^{\mathcal{G}_e} + \vartheta_{Z_1 X_2}^{\mathcal{G}_e} + 2\vartheta_{X_2 Z_3}^{\mathcal{G}_e} \right) - \frac{1}{4} \left(\vartheta_{Z_3}^{\mathcal{G}_e} + 2\vartheta_{X_4}^{\mathcal{G}_e} + 2\vartheta_{Z_5}^{\mathcal{G}_e} + \vartheta_{Z_3 X_4}^{\mathcal{G}_e} + 2\vartheta_{X_4 Z_5}^{\mathcal{G}_e} \right) \\
& + \frac{1}{4} \left(2\vartheta_{X_2}^{\mathcal{G}_o} + \vartheta_{Z_3}^{\mathcal{G}_o} + \vartheta_{X_2 Z_3}^{\mathcal{G}_o} \right) - \frac{1}{2} \left(2\vartheta_{X_4}^{\mathcal{G}_o} + \vartheta_{Z_5}^{\mathcal{G}_o} + \vartheta_{X_4 Z_5}^{\mathcal{G}_o} \right) - \frac{3}{4} \left(\vartheta_{Z_3}^{\mathcal{G}_e} + 2\vartheta_{X_4}^{\mathcal{G}_e} + \vartheta_{Z_3 X_4}^{\mathcal{G}_e} \right) + \left(\vartheta_{Z_3}^{\mathcal{G}_e} \right) \\
& = \vartheta_{X_2}^{\mathcal{G}_o} - \vartheta_{X_4}^{\mathcal{G}_e},
\end{aligned} \tag{F8}$$

where among the added terms, the first three terms are from Eq. (85) and the rest are from Eq. (84). In this way, we can change the basis in Eq. (86) into for $k = 1, \dots, \frac{n}{2}$,

$$\left\{ -\vartheta_{X_{2k-1}}^{\mathcal{G}_e} + \vartheta_{X_{2k+1}}^{\mathcal{G}_o}, \vartheta_{X_{2k}}^{\mathcal{G}_o} - \vartheta_{X_{2k+2}}^{\mathcal{G}_e} \right\}. \tag{F9}$$

The change from Eqs. (84), (85), (86) to Eqs. (84), (85), (F9) is obviously invertible, so now we prove that the latter is a basis for L_R^G .

In fact, Eqs. (84), (85), (F9) is a valid output of Algorithm 3. Initially, $A = \{\vartheta_a^{\mathcal{G}_e}, \vartheta_a^{\mathcal{G}_o} : a \sim \Omega\}$. We then put cycles from Eq. (85) into C one by one (the order does not matter). For example, when we pick $c = \mathcal{Q}^\top \left(e_a^{\mathcal{G}_e} + e_{a'}^{\mathcal{G}_e} \right)$ ($a \in \mathbf{P}^{2k, 2k+1}$, $w(a) = 2$, $a' = \mathcal{G}_e(a)$) into C , we choose $\vartheta_a^{\mathcal{G}_e}$ as ϑ and remove it from A . The case when $\mathcal{G}_e$ is replaced by $\mathcal{G}_o$ is similar. These ϑ removed does not appear in other cycles in Eq. (85), so the c picked each time indeed belongs to $\widetilde{L}_R^G \cap \text{span}(A)$ at the time.

Now C contains elements from Eq. (85), and A can be partitioned into disjoint sets: $A = \bigcup_{k=1, \dots, n/2} \{\vartheta_a^{\mathcal{G}_e} : a \in \mathbf{P}^{2k-1, 2k}\} \cup \bigcup_{k=1, \dots, n/2} \{\vartheta_a^{\mathcal{G}_e} : a \in \mathbf{P}^{2k-1, 2k}\}$. This partition corresponds to the partition in Eq. (84), and we put them into C one by one. For $k = 1, \dots, n/2$, first pick

$$c = \mathcal{Q}^\top \left(e_{X_{2k-1} X_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{Y_{2k-1} Y_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{X_{2k-1} Y_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{Y_{2k-1} X_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{Z_{2k-1} Z_{2k}}^{\mathcal{G}_e} \right), \tag{F10}$$

which corresponds to picking

$$\vartheta = \vartheta_{X_{2k-1} X_{2k}}^{\mathcal{G}_e}, \vartheta_{Y_{2k-1} Y_{2k}}^{\mathcal{G}_e}, \vartheta_{X_{2k-1} Y_{2k}}^{\mathcal{G}_e}, \vartheta_{Y_{2k-1} X_{2k}}^{\mathcal{G}_e}, \vartheta_{Z_{2k-1} Z_{2k}}^{\mathcal{G}_e}. \tag{F11}$$

Then pick in the order of

$$\begin{aligned}
c = & \mathcal{Q}^\top \left(e_{Y_{2k-1} Z_{2k}}^{\mathcal{G}_e} + e_{Y_{2k-1} I_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{Y_{2k-1} I_{2k}}^{\mathcal{G}_e} + e_{X_{2k-1} Z_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{X_{2k-1} Z_{2k}}^{\mathcal{G}_e} + e_{X_{2k-1} I_{2k}}^{\mathcal{G}_e} \right), \\
& \mathcal{Q}^\top \left(e_{Z_{2k-1} Y_{2k}}^{\mathcal{G}_e} + e_{I_{2k-1} Y_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{I_{2k-1} Y_{2k}}^{\mathcal{G}_e} + e_{Z_{2k-1} X_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(e_{Z_{2k-1} X_{2k}}^{\mathcal{G}_e} + e_{I_{2k-1} X_{2k}}^{\mathcal{G}_e} \right),
\end{aligned} \tag{F12}$$

which corresponds to picking

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_{Y_{2k-1}Z_{2k}}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{Y_{2k-1}I_{2k}}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{X_{2k-1}Z_{2k}}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{Z_{2k-1}Y_{2k}}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{I_{2k-1}Y_{2k}}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{Z_{2k-1}X_{2k}}^{\mathcal{G}_e}. \quad (\text{F13})$$

Finally pick

$$\boldsymbol{c} = \mathcal{Q}^\top \left(\boldsymbol{e}_{Z_{2k-1}I_{2k}}^{\mathcal{G}_e} \right), \mathcal{Q}^\top \left(\boldsymbol{e}_{I_{2k-1}Z_{2k}}^{\mathcal{G}_e} \right), \quad (\text{F14})$$

which corresponds to picking

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_{Z_{2k-1}I_{2k}}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{I_{2k-1}Z_{2k}}^{\mathcal{G}_e}. \quad (\text{F15})$$

And similarly for replacing $\mathcal{G}_e$ by $\mathcal{G}_o$. One can verify that in this order each time $\boldsymbol{c} \in \widetilde{L}_R^G \cap A$, and after picking cycles in Eq. (84) in this order, $A = \{\boldsymbol{\vartheta}_{X_k}^{\mathcal{G}_e}, \boldsymbol{\vartheta}_{X_k}^{\mathcal{G}_o} : k \in [n]\}$.

Then put elements in Eq. (F9) into C one by one (order does not matter). When $\boldsymbol{c} = -\boldsymbol{\vartheta}_{X_{2k-1}}^{\mathcal{G}_e} + \boldsymbol{\vartheta}_{X_{2k+1}}^{\mathcal{G}_o}$, pick $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_{X_{2k+1}}^{\mathcal{G}_o}$, and when $\boldsymbol{c} = \boldsymbol{\vartheta}_{X_{2k}}^{\mathcal{G}_o} - \boldsymbol{\vartheta}_{X_{2k+2}}^{\mathcal{G}_e}$ is picked, pick $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_{X_{2k+2}}^{\mathcal{G}_e}$.

At this stage C contains all elements from Eqs. (84), (85), (F9) and $A = \{\boldsymbol{\vartheta}_{X_k}^{\mathcal{G}_e} : k \in [n]\}$. It remains to prove that $\widetilde{L}_R^G \cap \text{span}(\{\boldsymbol{\vartheta}_{X_k}^{\mathcal{G}_e} : k \in [n]\}) = \{\mathbf{0}\}$. We have

$$\sum_{k=1}^{\frac{n}{2}} a_{2k-1} \boldsymbol{\vartheta}_{X_{2k-1}}^{\mathcal{G}_e} + a_{2k} \boldsymbol{\vartheta}_{X_{2k}}^{\mathcal{G}_e} \in L_R \Leftrightarrow \sum_{k=1}^{\frac{n}{2}} -a_{2k-1} \boldsymbol{\vartheta}_{\{2k\}}^M - a_{2k} \boldsymbol{\vartheta}_{\{2k-1\}}^M \in L_R \Leftrightarrow \forall k, a_k = 0 \quad (\text{F16})$$

where the first equivalence is because

$$\begin{aligned} & \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k-1\}}^S + \boldsymbol{e}_{X_{2k-1}}^{\mathcal{G}_e} + \boldsymbol{e}_{\{2k-1,2k\}}^M \right) - \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k-1\}}^S + \boldsymbol{e}_{\{2k-1\}}^M \right) + \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k-2\}}^S + \boldsymbol{e}_{X_{2k-2}}^{\mathcal{G}_o} + \boldsymbol{e}_{\{2k-2,2k-1\}}^M \right) \\ & - \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k-2,2k+1\}}^S + \boldsymbol{e}_{X_{2k-2}X_{2k+1}}^{\mathcal{G}_o} + \boldsymbol{e}_{\{2k-2,2k-1,2k,2k+1\}}^M \right) + \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k+1\}}^S + \boldsymbol{e}_{X_{2k+1}}^{\mathcal{G}_o} + \boldsymbol{e}_{\{2k,2k+1\}}^M \right) \\ & = \boldsymbol{\vartheta}_{X_{2k-1}}^{\mathcal{G}_e} + \boldsymbol{\vartheta}_{\{2k\}}^M \in L_R \end{aligned} \quad (\text{F17})$$

and similarly $\boldsymbol{\vartheta}_{X_{2k}}^{\mathcal{G}_e} + \boldsymbol{\vartheta}_{\{2k-1\}}^M \in L_R$. The second equivalence is because $\boldsymbol{\vartheta}_{\{k\}}^M$ changes independently under the single-qubit depolarizing gauges. Hence now $\widetilde{L}_R^G \cap \text{span}(A) = \emptyset$ and Algorithm 3 terminates. By the correctness of Algorithm 3, C , which contains Eqs. (84), (85), (F9) is a basis for $\widetilde{L}_R^G$ and thus a basis for L_R^G . Thus Eqs. (84), (85), (86) is a basis for L_R^G .

Next, we give a way to supplement Eqs. (84), (85), (86) into a reduced cycle basis for L_R . We claim the following reduced cycles work:

$$\begin{aligned} & \left\{ \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k-1\}}^S + \boldsymbol{e}_{X_{2k-1}}^{\mathcal{G}_e} + \boldsymbol{e}_{\{2k-1,2k\}}^M \right), \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k\}}^S + \boldsymbol{e}_{X_{2k}}^{\mathcal{G}_e} + \boldsymbol{e}_{\{2k-1,2k\}}^M \right) : k = 1, \dots, \frac{n}{2} \right\} \\ & \setminus \left\{ \mathcal{Q}^\top \left(\boldsymbol{e}_{\{1\}}^S + \boldsymbol{e}_{X_1}^{\mathcal{G}_e} + \boldsymbol{e}_{\{1,2\}}^M \right) \right\} \end{aligned} \quad (\text{F18})$$

$$\left\{ \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k\}}^S + \boldsymbol{e}_{X_{2k}}^{\mathcal{G}_o} + \boldsymbol{e}_{\{2k,2k+1\}}^M \right), \mathcal{Q}^\top \left(\boldsymbol{e}_{\{2k+1\}}^S + \boldsymbol{e}_{X_{2k+1}}^{\mathcal{G}_o} + \boldsymbol{e}_{\{2k,2k+1\}}^M \right) : k = 1, \dots, \frac{n}{2} \right\} \quad (\text{F19})$$

$$\left\{ \mathcal{Q}^\top \left(\boldsymbol{e}_{\{k,k+1\}}^S + \boldsymbol{e}_{\{k,k+1\}}^M \right) : k \in [n] \right\} \quad (\text{F20})$$

$$\left\{ \mathcal{Q}^\top \left(\boldsymbol{e}_{\{k\}}^S + \boldsymbol{e}_{\{k\}}^M \right) : k \in [n] \right\} \quad (\text{F21})$$

$$\left\{ \mathcal{Q}^\top \left(\boldsymbol{e}_{\{n,3\}}^S + \boldsymbol{e}_{X_n X_3}^{\mathcal{G}_o} + \boldsymbol{e}_{\{n,1,2,3\}}^M \right) \right\}. \quad (\text{F22})$$

As a sanity check, these give $(n-1) + n + n + n + 1 = 4n$ dimensions, which together with the $23n$ dimensions from L_R^G gives the $27n$ dimensions for L_R .

To prove that Eqs. (84), (85), (86), (F18), (F19), (F20), (F21), (F22) form a basis of L_R , again we may first perform an invertible linear transformation on them. We change Eq. (F22) into

$$\mathcal{Q}^\top \left(e_{\{n,3\}}^S + e_{X_n X_3}^{\mathcal{G}_o} + e_{\{n,1,2,3\}}^M \right) - \mathcal{Q}^\top \left(e_{\{n\}}^S + e_{X_n}^{\mathcal{G}_o} + e_{\{n,1\}}^M \right) - \mathcal{Q}^\top \left(e_{\{3\}}^S + e_{X_3}^{\mathcal{G}_o} + e_{\{2,3\}}^M \right) = \vartheta_{\{1,2\}}^M. \quad (\text{F23})$$

We then change Eq. (86) into Eq. (F9) and then change for example $\vartheta_{X_{2k}}^{\mathcal{G}_o} - \vartheta_{X_{2k+2}}^{\mathcal{G}_e}$ into

$$\begin{aligned} & \vartheta_{X_{2k}}^{\mathcal{G}_o} - \vartheta_{X_{2k+2}}^{\mathcal{G}_e} - \mathcal{Q}^\top \left(e_{\{2k\}}^S + e_{X_{2k}}^{\mathcal{G}_o} + e_{\{2k,2k+1\}}^M \right) + \mathcal{Q}^\top \left(e_{\{2k+2\}}^S + e_{X_{2k+2}}^{\mathcal{G}_e} + e_{\{2k+1,2k+2\}}^M \right) \\ & + \mathcal{Q}^\top \left(e_{\{2k\}}^S + e_{\{2k\}}^M \right) - \mathcal{Q}^\top \left(e_{\{2k+2\}}^S + e_{\{2k+2\}}^M \right) \\ & = \vartheta_{\{2k+1,2k+2\}}^M - \vartheta_{\{2k,2k+1\}}^M, \end{aligned} \quad (\text{F24})$$

and similarly for the $-\vartheta_{X_{2k-1}}^{\mathcal{G}_e} + \vartheta_{X_{2k+1}}^{\mathcal{G}_o}$ s, except that $-\vartheta_{X_1}^{\mathcal{G}_e} + \vartheta_{X_3}^{\mathcal{G}_o}$ is changed into

$$-\vartheta_{X_1}^{\mathcal{G}_e} + \vartheta_{X_3}^{\mathcal{G}_o} - \mathcal{Q}^\top \left(e_{\{3\}}^S + e_{X_3}^{\mathcal{G}_o} + e_{\{2,3\}}^M \right) + \mathcal{Q}^\top \left(e_{\{3\}}^S + e_{\{3\}}^M \right) = -\vartheta_{X_1}^{\mathcal{G}_e} - \vartheta_{\{2\}}^M - \vartheta_{\{2,3\}}^M. \quad (\text{F25})$$

That is, Eq. (86) is changed into

$$\{\vartheta_{\{k,k+1\}}^M - \vartheta_{\{k+1,k+2\}}^M : k \in [n], k \neq 1\} \quad (\text{F26})$$

together with Eq. (F25).

Hence now it suffices to prove that Eqs. (85), (84), (F18), (F19), (F25), (F20), (F21), (F26), (F23) form a basis for L_R . Again we resort to Algorithm 3. This time the algorithm tries to find a basis for L_R , and initially $A = \{\vartheta_v^S, \vartheta_v^M, \vartheta_a^{\mathcal{G}_e}, \vartheta_a^{\mathcal{G}_o} : v \in \Omega, a \sim \Omega\}$. We first pick elements in Eqs. (85), (84) into C , using the same procedure described previously. Now $A = \{\vartheta_v^S, \vartheta_v^M, \vartheta_{X_k}^{\mathcal{G}_e}, \vartheta_{X_k}^{\mathcal{G}_o} : v \in \Omega, k \in [n]\}$.

Then put elements in Eqs. (F18), (F19) into C . For example, when $\mathcal{Q}^\top \left(e_{\{2k-1\}}^S + e_{X_{2k-1}}^{\mathcal{G}_e} + e_{\{2k-1,2k\}}^M \right)$ is picked, choose $\vartheta = \vartheta_{X_{2k-1}}^{\mathcal{G}_e}$. Afterwards, $A = \{\vartheta_v^S, \vartheta_v^M : v \in \Omega\} \cup \{\vartheta_{X_1}^{\mathcal{G}_e}\}$.

Next, put Eq. (F25) into C , picking $\vartheta = \vartheta_{X_1}^{\mathcal{G}_e}$. Now $A = \{\vartheta_v^S, \vartheta_v^M : v \in \Omega\}$.

Then put elements in Eq. (F20) into C , and when $\mathcal{Q}^\top \left(e_{\{k,k+1\}}^S + e_{\{k,k+1\}}^M \right)$ is picked, set ϑ to be $\vartheta_{\{k,k+1\}}^S$. What are left in A at this time are $A = \{\vartheta_{\{k\}}^S : k \in [n]\} \cup \{\vartheta_v^M : v \sim \Omega\}$.

Now put elements in Eq. (F21) into C , this time when $\mathcal{Q}^\top \left(e_{\{k\}}^S + e_{\{k\}}^M \right)$ is picked, let $\vartheta = \vartheta_{\{k\}}^S$. At this time $A = \{\vartheta_v^M : v \sim \Omega\}$.

Then put elements in Eq. (F26) as c and put into C , in the increasing order of k . When $\vartheta_{\{k,k+1\}}^M - \vartheta_{\{k+1,k+2\}}^M$ is picked, use $\vartheta = \vartheta_{\{k,k+1\}}^M$. Now A is left with $\{\vartheta_{\{k\}}^M : k \in [n]\} \cup \{\vartheta_{\{1,2\}}^M\}$.

Finally put Eq. (F23) into C and let $\vartheta = \vartheta_{\{1,2\}}^M$. A is now $\{\vartheta_{\{k\}}^M : k \in [n]\}$. Since each $\vartheta_{\{k\}}^M$ changes independently under single-qubit depolarization gauges, we conclude that $\text{span}(A) \cap L_R = \emptyset$ and Algorithm 3 terminates. By the correctness of the Algorithm 3, by this time C , which contains Eqs. (85), (84), (F18), (F19), (F25), (F20), (F21), (F26), (F23), form a basis for L_R . Thus Eqs. (84), (85), (86), (F18), (F19), (F20), (F21), (F22) form a basis for L_R .

-
- [1] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM Rev.* **41**, 303 (1999).
 - [2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, NY, USA, 2010).
 - [3] N. Gisin and R. Thew, Quantum communication, *Nat. Photonics* **1**, 165 (2007).
 - [4] H. J. Kimble, The quantum internet, *Nature* **453**, 1023 (2008).
 - [5] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
 - [6] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photonics* **5**, 222 (2011).
 - [7] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).

- [8] T. Proctor, K. Young, A. D. Baczewski, and R. Blume-Kohout, Benchmarking quantum computers, [arXiv:2407.08828](#).
- [9] A. Hashim, L. B. Nguyen, N. Goss, B. Marinelli, R. K. Naik, T. Chistolini, J. Hines, J. P. Marceaux, Y. Kim, P. Gokhale, *et al.*, Practical introduction to benchmarking and characterization of quantum computers, *PRX Quantum* **6**, 030202 (2025).
- [10] S. Flammia and J. Wallman, Efficient estimation of Pauli channels, *ACM Trans. Quantum Comput.* **1**, 1 (2020).
- [11] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, *Nat. Commun.* **10**, 5347 (2019).
- [12] S. T. Flammia and R. O'Donnell, Pauli error estimation via population recovery, *Quantum* **5**, 549 (2021).
- [13] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, *Nat. Phys.* **16**, 1184 (2020).
- [14] S. T. Flammia, in *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, Leibniz International Proceedings in Informatics (LIPIcs), edited by F. Le Gall and T. Morimae (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022), Vol. 232, pp. 4:1–4:10.
- [15] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Pauli channels can be estimated from syndrome measurements in quantum error correction, *Quantum* **6**, 809 (2022).
- [16] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Learning logical Pauli noise in quantum error correction, *Phys. Rev. Lett.* **130**, 200601 (2023).
- [17] E. Van Den Berg, Z. K. Mineev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors, *Nat. Phys.* **19**, 1116 (2023).
- [18] S. Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, and L. Jiang, The learnability of Pauli noise, *Nat. Commun.* **14**, 52 (2023).
- [19] S. Chen, S. Zhou, A. Seif, and L. Jiang, Quantum advantages for Pauli channel estimation, *Phys. Rev. A* **105**, 032435 (2022).
- [20] A. Carignan-Dugas, D. Dahlen, I. Hincks, E. Ospadov, S. J. Beale, S. Ferracin, J. Skanes-Norman, J. Emerson, and J. J. Wallman, The error reconstruction and compiled calibration of quantum computing cycles, [arXiv:2303.17714](#).
- [21] E. Van den Berg and P. Wocjan, Techniques for learning sparse Pauli–Lindblad noise models, *Quantum* **8**, 1556 (2024).
- [22] E. T. Hockings, A. C. Doherty, and R. Harper, Scalable noise characterization of syndrome-extraction circuits with averaged circuit eigenvalue sampling, *PRX Quantum* **6**, 010334 (2025).
- [23] E. H. Chen, T. J. Yoder, Y. Kim, N. Sundaresan, S. Srinivasan, M. Li, A. D. Córcoles, A. W. Cross, and M. Takita, Calibrated decoders for experimental quantum error correction, *Phys. Rev. Lett.* **128**, 110504 (2022).
- [24] E. Knill, Quantum computing with realistically noisy devices, *Nature* **434**, 39 (2005).
- [25] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [26] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, *et al.*, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, *Phys. Rev. X* **11**, 041039 (2021).
- [27] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
- [28] H.-Y. Huang, S. T. Flammia, and J. Preskill, Foundations for learning from noisy quantum experiments, [arXiv:2204.13691](#).
- [29] One may wonder why we only consider *linear* functions of the logarithmic Pauli eigenvalues. This is because the results of any experiments can be determined by some learnable linear functions, so they suffice to characterize the learnability. See Theorem 6 for more details.
- [30] B. Bollobás, *Modern Graph Theory* (Springer Science & Business Media, New York, NY, USA, 1998), Vol. 184.
- [31] P. M. Gleiss, J. Leydold, and P. F. Stadler, Circuit bases of strongly connected digraphs, *Discuss. Math. Graph Theory* **23**, 241 (2003).
- [32] Rigorously speaking, L is a subspace of the dual space X' . Here we identify X with X' using the standard basis for defining X .
- [33] If $\mathcal{Q}$ is not injective, the reduced model must contain redundant parameters. One can thus redefine the reduced model to eliminate the redundancy and retrieve an injective embedding map.
- [34] R. Harper, I. Hincks, C. Ferrie, S. T. Flammia, and J. J. Wallman, Statistical analysis of randomized benchmarking, *Phys. Rev. A* **99**, 052350 (2019).
- [35] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of addressability by simultaneous randomized benchmarking, *Phys. Rev. Lett.* **109**, 240504 (2012).
- [36] E. Pelaez Cisneros, V. Omole, P. Gokhale, R. Rines, K. N. Smith, M. A. Perlin, and A. Hashim, Average circuit eigenvalue sampling on NISQ devices, [arXiv:2403.12857](#).
- [37] IBM Quantum, <https://quantum-computing.ibm.com> (2022).
- [38] Y. Mao and F. R. Kschischang, On factor graphs and the Fourier transform, *IEEE Trans. Inf. Theory* **51**, 1635 (2005).
- [39] A Python implementation of gate set Pauli noise learning protocol in the example of parallel CZ on a ring with nearest-neighbor 2-local noise, 2025, <https://github.com/csenrui/PauliGST>
- [40] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, [arXiv:2405.08810](#).
- [41] E. P. Cisneros, V. Omole, P. Gokhale, R. Rines, K. N. Smith, M. A. Perlin, and A. Hashim, Average circuit eigenvalue sampling on NISQ devices, [arXiv:2403.12857](#).
- [42] C. Ostrove, K. Rudinger, S. Seritan, K. Young, and R. Blume-Kohout, Near-Minimal Gate Set Tomography Experiment Designs, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Bellevue, WA, USA, 2023), Vol. 1, pp. 1422–1432.

- [43] S. Chen, C. Oh, S. Zhou, H.-Y. Huang, and L. Jiang, Tight bounds on Pauli channel learning without entanglement, *Phys. Rev. Lett.* **132**, 180805 (2024).
- [44] O. Fawzi, A. Oufkir, and D. S. França, Lower bounds on learning Pauli channels with individual measurements, *IEEE Trans. Inf. Theory* **71**, 2642 (2025).
- [45] S. Chen and W. Gong, Efficient Pauli channel estimation with logarithmic quantum memory, *PRX Quantum* **6**, 020323 (2025).
- [46] X. D. Trinh and N. Yu, Adaptivity is not helpful for Pauli channel learning, *Quantum* **9**, 1864 (2025).
- [47] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [48] S. Ferracin, A. Hashim, J.-L. Ville, R. Naik, A. Carignan-Dugas, H. Qassim, A. Morvan, D. I. Santiago, I. Siddiqi, and J. J. Wallman, Efficiently improving the performance of noisy quantum computers, *Quantum* **8**, 1410 (2024).
- [49] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, *et al.*, Evidence for the utility of quantum computing before fault tolerance, *Nature* **618**, 500 (2023).
- [50] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [51] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, *Phys. Rev. X* **8**, 031027 (2018).
- [52] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the NISQ to the fault-tolerant quantum computing eras, *PRX Quantum* **3**, 010345 (2022).
- [53] E. H. Chen, S. Chen, L. E. Fischer, A. Eddins, L. C. Govia, B. Mitchell, A. He, Y. Kim, L. Jiang, and A. Seif, Disambiguating Pauli noise in quantum computers, *arXiv:2505.22629*.
- [54] E. Nielsen, K. Young, and R. Blume-Kohout, First-order gauge-invariant error rates in quantum processors, in *APS March Meeting Abstracts* (American Physical Society, Chicago, 2022), Vol. 2022, p. M38-009.
- [55] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, A taxonomy of small Markovian errors, *PRX Quantum* **3**, 020335 (2022).
- [56] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, *et al.*, Realization of real-time fault-tolerant quantum error correction, *Phys. Rev. X* **11**, 041058 (2021).
- [57] K. Singh, C. E. Bradley, S. Anand, V. Ramesh, R. White, and H. Bernien, Mid-circuit correction of correlated phase errors using an array of spectator qubits, *Science* **380**, 1265 (2023).
- [58] I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, *Nature* **614**, 676 (2023).
- [59] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, *Nature* **626**, 58 (2024).
- [60] S. J. Beale and J. J. Wallman, Randomized compiling for subsystem measurements, *arXiv:2304.06599*.
- [61] Z. Zhang, S. Chen, Y. Liu, and L. Jiang, Generalized cycle benchmarking algorithm for characterizing midcircuit measurements, *PRX Quantum* **6**, 010310 (2025).
- [62] J. Hines and T. Proctor, Pauli noise learning for mid-circuit measurements, *Phys. Rev. Lett.* **134**, 020602 (2025).
- [63] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia, Logical randomized benchmarking, *arXiv:1702.03688*.
- [64] C. Rouzé and D. S. França, Efficient learning of the structure and parameters of local Pauli noise channels, *arXiv:2307.02959*.
- [65] C. Hsiao, Identification, *Handb. Econom.* **1**, 223 (1983).
- [66] In the literature of algebraic graph theory, people sometimes allow reverse edges (i.e., $e_{ik} = (v_{ik}, v_{ik-1})$) in the definition of a cycle and refer to our definition as an *oriented cycle* or a *circuit*. Since all graphs considered in this work are either strongly-connected or a union of strongly-connected components, and it has been shown that oriented cycles can linearly represent any cycle for such graphs [31], working with our current definition would not cause any inconsistency. Also, we allow a cycle to contain the same edge multiple times, which is referred to as a *closed walk* in standard terminology.
- [67] Although ρ_a is not uniquely defined when P_a contains identity, choosing any ρ_a that satisfies this condition will make no observable difference.
- [68] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, *Phys. Rev. A* **77**, 032322 (2008).
- [69] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, *J. Opt. B* **7**, S347 (2005).
- [70] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
- [71] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [72] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [73] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [74] Though $\mathcal{Q}$ is not invertible, it is injective and we know that $\forall |s| = 1, \mathfrak{d}_s \in \text{Im } \mathcal{Q}$, thus $\mathcal{Q}^{-1}(\mathfrak{d}_s)$ is uniquely defined.
- [75] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, *Phys. Rev. A* **103**, 042605 (2021).
- [76] A. G. Fowler and J. M. Martinis, Quantifying the effects of local many-qubit errors and nonlocal two-qubit errors on the surface code, *Phys. Rev. A* **89**, 032316 (2014).
- [77] N. H. Nickerson and B. J. Brown, Analysing correlated noise on the surface code using adaptive decoding algorithms, *Quantum* **3**, 131 (2019).
- [78] P. Clifford, Markov random fields in statistics, in *Disorder in physical systems: A volume in honour of John M. Hammersley* (Clarendon Press, Oxford, 1990), p. 19.

- [79] Note that, we do not require those Pauli diagonal maps to be completely positive. Similarly, when we say our quasi-local model is equivalent to the sparse Pauli-Lindblad model, we have not required the positivity of the model parameters.
- [80] H.-Y. Hu, A. Gu, S. Majumder, H. Ren, Y. Zhang, D. S. Wang, Y.-Z. You, Z. Mineev, S. F. Yelin, and A. Seif, Demonstration of robust and efficient quantum property learning with shallow shadows, *Nat. Commun.* **16**, 2943 (2025).
- [81] Note that the CNOT and the echoed cross-resonance (ECR) gates are all equivalent to CZ up to single-qubit rotations. Our theory can similarly be applied to them.
- [82] As a simpler proof that the dimension of L_R^G equals $23n$, first note that the gauge space has dimension n , which means the dimension of L_R^G is at least $24n - n = 23n$. Then, around Eqs. (F16) we show that L_R^G has trivial intersection with an n -dimensional subspace within X_R^G , meaning its dimension is no larger than $24n - n = 23n$. Putting these two together, we conclude that the dimension of L_R^G is exactly $23n$.
- [83] A. Calzona, M. Papič, P. Figueroa-Romero, and A. Auer, Multi-layer cycle benchmarking for high-accuracy error characterization, *arXiv:2412.09332*.
- [84] <https://github.com/csenrui/PauliGST>.
- [85] C. Berge, *The Theory of Graphs* (Courier Corporation, New York, NY, USA, 2001).
- [86] D. Grier and L. Schaeffer, The classification of Clifford gates over qubits, *Quantum* **6**, 734 (2022).