

Disambiguating Pauli Noise in Quantum Computers

Edward H. Chen^{1,*}, Senrui Chen^{2,†,‡}, Laurin E. Fischer^{3,4,§}, Andrew Eddins⁵, Luke C. G. Govia⁵,
Brad Mitchell⁵, Andre He⁶, Youngseok Kim⁶, Liang Jiang^{2,||}, and Alireza Seif^{6,¶}

¹*IBM Research, Research Triangle Park, North Carolina, USA*

²*Pritzker School of Molecular Engineering, University of Chicago, Chicago 60637, USA*

³*IBM Research, Rüschlikon, Switzerland*

⁴*Theory and Simulation of Materials, École Polytechnique Fédérale de Lausanne,
1015 Lausanne, Switzerland*

⁵*IBM Research, San Jose, California, USA*

⁶*IBM Research, Yorktown Heights, New York, USA*

 (Received 26 June 2025; revised 24 March 2026; accepted 9 July 2026; published 3 September 2026)

To successfully perform quantum computations, it is often necessary to first accurately characterize the noise in the underlying hardware. However, it is well known that fundamental limitations prevent the unique identification of the noise model. This raises the question of whether these limitations impact the ability to predict noisy dynamics and mitigate errors. Here, we show, both theoretically and experimentally, that when learnable parameters are self-consistently characterized, the *unlearnable* (gauge) degrees of freedom do not impact predictions of noisy dynamics or error mitigation. We use the recently introduced framework of gate set Pauli noise learning to efficiently and self-consistently characterize and mitigate noise of a complete gate set, including state preparation, measurements, single-qubit gates and multi-qubit entangling Clifford gates. We validate our approach through experiments with up to 92 qubits and show that, while the gauge choice does not affect error-mitigated observables, optimizing it reduces sampling overhead.

DOI: 10.1103/69wc-gz16

I. INTRODUCTION

Quantum computers are believed to be exponentially faster than classical computers for many important problems [1]. However, noise limits the performance of the quantum hardware, motivating the widespread efforts to characterize the noise in order to address it. Important areas where noise learning protocols are expected to have on-going impact include: quantifying improvements to hardware architectures [2], mitigating the impact of noise on observables with additional quantum and classical processing [3], or improving algorithms needed to actively correct noise soon after it occurs [4]. As progress is made

along all directions, it is increasingly accepted that quantum computations will also *continuously*, as opposed to abruptly, improve in accuracy [5,6].

Recent progress toward building larger quantum computers has highlighted a need for scalable methods to fully characterize all possible types of quantum noise, which can be as intractable to classically model as the quantum algorithm being executed [7]. To reduce the complexity of the learning task, the predominant sources of noise are assumed to only impact the qubit subspace and are also physically localized to neighboring qubits on the device. Upon transforming the underlying noise using randomized compiling or Pauli twirling [8], a Pauli noise model becomes a practical choice because it can be made as complex as necessary while remaining classically tractable [5]. In fact, it was recently shown that a learned noise model could be used to effectively mitigate noise in applications which require accurate estimates of expectation values [3,9,10].

Such error mitigation strategies can in principle yield unbiased estimators at the cost of additional quantum circuit executions (shots), with the assumption that the device noise is faithfully captured by the learned noise model [11]. The original proposal of Ref. [12] addressed this requirement by building on gate set tomography (GST) [13], which learns gate, state-preparation, and measurement noise self-consistently and naturally accounts for gauge degrees of

*Contact author: ehchen@alum.mit.edu

†Contact author: csenrui@gmail.com

‡Present address: California Institute of Technology, Pasadena, CA, USA.

§Contact author: aur@zurich.ibm.com

||Contact author: liang.jiang@uchicago.edu

¶Contact author: alireza.seif@ibm.com

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International* license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

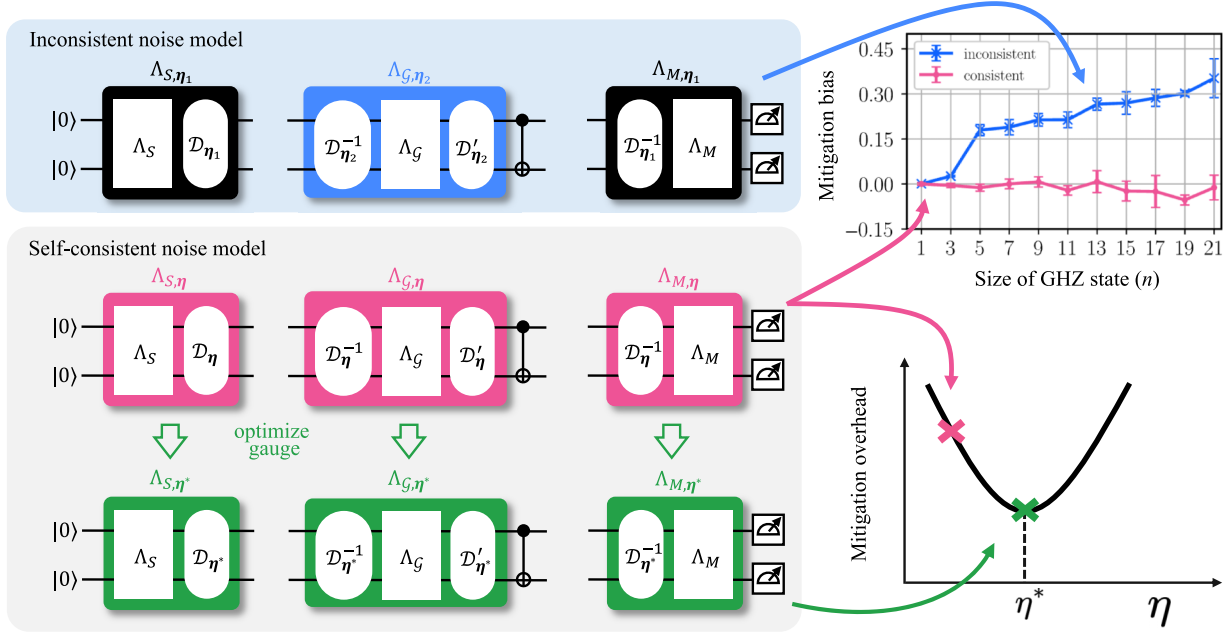


FIG. 1. Overview of results. Leading error mitigation methods based on Pauli noise models presuppose accurate knowledge of all the hardware error rates. This foundational assumption, however, is false, as it has been proven that such a noise model cannot be uniquely determined by experiments, even in principle. Without accounting for this indeterminacy, previous experiments implicitly used an *inconsistent* set of the gauge parameters $\{\mathcal{D}_{\eta_1}, \mathcal{D}_{\eta_2}\}$ across the quantum gate set, e.g., different gauge choices for state preparation and measurement and the two-qubit gate (top, blue). We show that a self-consistent set of gauge parameters (middle, pink) is necessary for unbiased quantum error mitigation, as exemplified here in the mitigation bias of state preparation experiment of an $n = 21$ entangled state known as the Greenberger-Horne-Zeilinger state (GHZ) state (upper right; details in Fig. 6). Furthermore, the choice of a consistent gauge can be optimized (bottom, green) to reduce the sampling overhead of error mitigation (bottom, right).

freedom in the noise description. However, error mitigation with GST has been limited in practice to one- and two-qubit systems due to its unfavorable scaling [14,15], with even three-qubit GST requiring considerable experimental effort [16]. Therefore, subsequent work turned to more scalable noise-learning approaches such as cycle benchmarking [17,18] and sparse Pauli-Lindblad models [3,9,19] to enable error mitigation on processors with tens to over a hundred qubits. These scalable methods achieve their efficiency by characterizing gate noise using SPAM-robust techniques and treating SPAM errors separately. While practically convenient, this separation comes at a cost. It has recently been shown that certain combinations of gate and SPAM noise parameters cannot be uniquely identified when learned independently, giving rise to gauge degrees of freedom in the Pauli noise model [20,21]. When SPAM and gate noise are characterized separately, these gauge-ambiguous combinations cannot be correctly resolved, which leads to inconsistencies in the predictions of the error model (see Fig. 1).

Although the existence of such gauge ambiguities has been established theoretically [21], their practical consequences for error mitigation have not been investigated. It has remained an open question how large these inconsistencies are on real hardware and whether they meaningfully degrade mitigation performance. Here, we address both

questions. We prove theoretically that by self-consistently inferring all the *learnable* parameters in a Pauli noise model, including SPAM and gate noise together, it is possible to predict the outcomes of any noisy experiment and successfully perform error mitigation, even without resolving the fundamentally unlearnable gauge parameters. We further provide extensive experimental evidence, quantifying the magnitude of gauge-induced inconsistencies in practical error mitigation and demonstrating that self-consistent learning eliminates them. The key enabler is the Pauli gate set learning method of Ref. [21], which treats SPAM and gate noise within a unified framework similar in spirit to GST [13] but specialized to Pauli noise for scalability and efficiency. Our experiments use this framework to identify the learnable parameters of Pauli noise channels and to *unambiguously* and efficiently characterize them. We review this framework in Sec. II A and discuss how gauge degrees of freedom emerge as a result of SPAM errors. We then discuss the application of this framework to a quasi-local noise model in Sec. II B.

In Sec. III A, we show that such noise models naturally enable a self-consistent and unbiased error mitigation strategy. Specifically, we show that when probabilistic error cancellation (PEC) [9] is implemented with self-consistently learned noise models and applied to SPAM and gate errors, it produces unbiased estimates of

observables that do not depend on the gauge degrees of freedom. We further show, surprisingly, that despite not impacting observables, changing the gauge parameters does impact the overhead of required shots for error mitigation. Building on this insight, we propose and demonstrate a scalable method for identifying the gauge parameters needed to minimize this sampling overhead.

We demonstrate the learning framework and our theoretical results by performing several experiments with increasing complexity and show that it reduces the bias in mitigated expectation values compared to previous approaches. Specifically, we start from a simple two-qubit example in Sec. IV A and show that inconsistencies in handling gauge degrees of freedom in previous error mitigation techniques lead to errors in mitigated expectation values, whereas our method provides consistent and accurate estimates. Building on these results, we next consider expectation values of high-weight stabilizers of Greenberger-Horne-Zeilinger (GHZ) [22] states on up to 21 qubits in Sec. IV B. In these experiments, we do not impose locality on the error model, but instead rely on the stabilizer nature of the target state to simplify the mitigation by only learning a subset of error parameters. We again observe that while inconsistencies limit the accuracy of previously used methods, our method succeeds in producing correct estimates. To showcase the scalability of the learning protocol, we consider brickwork circuits on a ring of 92 qubits and learn the full quasi-local noise model in Sec. IV C. We consider 92 single-qubit observables and again observe that our method generally reduces bias compared to previous techniques. In Secs. IV A–IV B, the Clifford structure of the circuits allows us to construct proxy mitigated values without physically implementing PEC by dividing measured noisy expectation values by model predictions. Finally, in Sec. IV D we perform an end-to-end PEC experiments on non-Clifford circuits of 20 qubits which confirm that optimizing the gauge of a self-consistently learned model can significantly reduce the shot overhead of error mitigation. See overview of all the sections in Fig. 1.

II. MATHEMATICAL PRELIMINARIES

A. Modeling and learning a gate set

A Pauli channel on n qubits is a stochastic mixture of n -qubit Pauli operators $P_a \in \mathcal{P}^n = \{I, X, Y, Z\}^{\otimes n}$ described by a 4^n -dimensional probability distribution $\{p_a\}$, known as the Pauli error rates. One property of Pauli channels is that they transform any Pauli operator P_a to itself up to a prefactor $\lambda_a \in [-1, 1]$, known as the Pauli eigenvalues. Mathematically, a Pauli channel can be represented in the following two ways,

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

Both representations have $4^n - 1$ degrees of freedom, as the trace-preserving condition of quantum channels requires $\sum_a p_a = 1$, or equivalently $\lambda_{I^{\otimes n}} = 1$. For now, we consider Pauli channels that are completely general. We will discuss Pauli channels with efficient parameterization (e.g., quasi-local Pauli channels) in the next subsection.

Our work considers a “gate set” [13] comprised of state preparation, measurement, single-qubit gates, and entangling gates (See Fig. 1). Concretely, let the ideal initial state be $\rho_0 = |0\rangle\langle 0|^{\otimes n}$, the measurement be the projection onto the computational basis $\mathcal{M}_Z$, the entangling gates be a finite collection of Clifford gates $\{\mathcal{G}\}$, and the single-qubit gates be arbitrary $\{\mathcal{U} = \bigotimes_{i=1}^n \mathcal{U}_i\}$. In practice, the gate set is noisy. We use a Pauli noise model to describe the noisy gate set, where state preparation, measurement, and entangling gates are subject to Pauli noise channels,

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

We further assume that the single-qubit gates have negligible noise (which can be relaxed to gate-independent noise [8]), and that the SPAM noise channels Λ^S and Λ^M are generalized depolarizing channels, which are Pauli channels whose Pauli error rates only depend on the support of the corresponding Pauli operators, and thus contain only $2^n - 1$ degrees of freedom. These assumptions about the noise channels can be physically enforced using randomized compiling or Pauli twirling given reasonably good single-qubit control, as have been widely adopted and verified in the literature [8,9,23,24].

Before we discuss how to learn the noisy gate set, it is crucial to note that not every noise parameter is identifiable [20,21,25]. To see this, we highlight the fact that for any quantum circuit and any observable, the noisy expectation value takes the following form,

$$\begin{aligned} \langle \tilde{o} \rangle &= \sum_{a_0, \dots, a_{T+1} \in \mathcal{P}^n} c_a \lambda_{a_0}^S \lambda_{a_1}^{\mathcal{G}_1} \cdots \lambda_{a_T}^{\mathcal{G}_T} \lambda_{a_{T+1}}^M \\ &= \sum_{a_0, \dots, a_{T+1} \in \mathcal{P}^n} c_a \Gamma_a. \end{aligned} \quad (3)$$

Here, there are T layers of entangling Clifford gates $\mathcal{G}_1, \dots, \mathcal{G}_T$ in the circuits, possibly interleaved by single-qubit gates. $\{c_a\}$ are real numbers depending only on the ideal circuits and the observables, but not on the noise parameters. $\Gamma_a = \lambda_{a_0}^S \lambda_{a_1}^{\mathcal{G}_1} \cdots \lambda_{a_T}^{\mathcal{G}_T} \lambda_{a_{T+1}}^M$ is a product of Pauli eigenvalues known as a Pauli path [26].

Importantly, Γ_a cannot be arbitrary monomials of Pauli eigenvalues. The allowed set of Γ_a can be described using a directed graph called the pattern transfer graph [20,21], which describes how the gate set transform Pauli operators. The *pattern* of an n -qubit Pauli operator a is an n -bit string, denoted by $\text{pt}(a)$, whose i th bit is 1 if a_i is a nontrivial Pauli and 0 otherwise. For example, $\text{pt}(IXIYZ) = 01011$. See Fig. 2 for an example of a

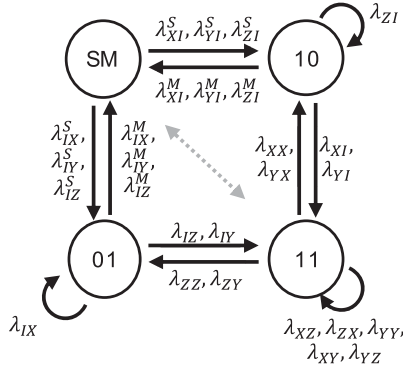


FIG. 2. Pattern transfer graph. Example of a 2-qubit gate set that contains CNOT as the only entangling gate, where any observable can be mapped to a cycle starting from the state-preparation and measurement (SM) node and back. The bi-directional gray dashed line has 18 SPAM eigenvalues associated with it that have been omitted for clarity. The numbers in the nodes denote *patterns* of Pauli operators, i.e., a bit string indicating which qubits have a nontrivial single-qubit Pauli operator supported on them.

pattern transfer graph for a 2-qubit system with controlled-NOT (CNOT) being the only entangling Clifford gate between a control qubit (left index) and a target qubit (right index), e.g. CNOT(IZ) $\rightarrow$ CNOT(ZZ). Each edge on the pattern transfer graph corresponds to a unique Pauli eigenvalue from one of the Pauli noise channels. Any path on the graph corresponds to a product of Pauli eigenvalues along the path. It is known that the set of valid $\{\Gamma_a\}$ has a one-to-one correspondence with the set of paths starting from and ending at the root node denoted as “SM” [20,21]. Consequently, the products of Pauli eigenvalues on cycles completely determine the outcomes of all possible experiments within the noisy gate set. If we transform the Pauli eigenvalues in a way that preserves the value of all cycles, then no experiments can witness such a transformation, which means there are gauge (i.e., non-identifiable) degrees of freedom. It was shown in Ref. [21] that all gauge transformations in the Pauli noise model can be expressed as

$$\Lambda^S \mapsto \Lambda_\eta^S = \mathcal{D}_\eta \circ \Lambda^S, \quad (4)$$

$$\Lambda^M \mapsto \Lambda_\eta^M = \Lambda^M \circ \mathcal{D}_\eta^{-1}, \quad (5)$$

$$\Lambda^G \mapsto \Lambda_\eta^G = \mathcal{D}'_\eta \circ \Lambda^G \circ \mathcal{D}_\eta^{-1}. \quad (6)$$

Here, $\mathcal{D}_\eta$ is any generalized depolarizing map, written as

$$\mathcal{D}_\eta(\rho) = \sum_{a \in \mathcal{P}^n} e^{-\eta_{\text{pt}(a)}} P_a \text{tr}(P_a \rho) / 2^n. \quad (7)$$

With a real vector η we refer to as the gauge parameters where $\eta_{0_n} = 0$ by the trace-preserving condition.

Recall that $\text{pt}(a)$ is the pattern of the the Pauli operator a .

$\mathcal{D}'_\eta$ is defined by $\mathcal{D}'_\eta = \mathcal{G}^{-1} \circ \mathcal{D}_\eta \circ \mathcal{G}$. Note that $\mathcal{D}_\eta$ commutes with any single-qubit gates. Thus, the transformations in Eq. (4) preserve any experimental outcome, and also all noise assumptions of the Pauli noise model. This is illustrated in Fig. 1. The remaining consideration is the positivity of the transformed channels—excluding Pauli channels on the boundary of the set of positive maps, any sufficiently small η yield physical channels [20]. Furthermore, in applications like error mitigation to be discussed later, it is acceptable to work with Λ_η that are not positive. Thus, a noisy gate set can be learned up to the $2^n - 1$ gauge parameters parameterized by η [21].

To learn the gate set self-consistently, we first define the logarithm of the Pauli eigenvalues $x_a = -\log \lambda_a$ for all the Pauli channels $\{\Lambda^S, \Lambda^M, \{\Lambda^G\}\}$. Let $\mathbf{x}$ be a vector comprised of all $\{x_a\}$, the length of which depends on the size of the gate set and the number of qubits. We will design a set of experiments to learn $\mathbf{x}$, where each experiment consists of a sequence of Clifford gates and a Pauli observable measured at the end, the expectation value of which satisfies $\langle \tilde{o} \rangle = \Gamma_a$. Taking the negative logarithm on both sides yields,

$$-\log \langle \tilde{o} \rangle = x_{a_0}^S + x_{a_1}^G + \dots + x_{a_r}^G + x_{a_{r+1}}^M, \quad (8)$$

which is a linear equation of $\mathbf{x}$. We emphasize that unlike Eq. (3) which holds for *any* observable for any general circuit, this expression requires only a monomial expression because it refers to a Pauli observable for a Clifford circuit. Combining the linear equations from all experiments, we arrive at

$$\mathbf{b} = F\mathbf{x}, \quad (9)$$

where $b_j = -\log \langle \tilde{o}_j \rangle$ is the (log) expectation value for the j -th measured Pauli observable on the j -th circuit, and F is called the design matrix. Our first goal is to collect enough experiments such that F has the maximal possible rank. That is, the dimension of the null space of F equals the number of gauge parameters, $2^n - 1$. This can be achieved by including some experiments that contain no entangling gates (called depth-0 experiments, $b = x_{a_0}^S + x_{a_1}^M$) or contain one layer of entangling gates (called depth-1 experiments, $b = x_{a_0}^S + x_{a_1}^G + x_{a_2}^M$); to improve the estimated precision of model parameters, we can also include experiments that concatenate multiple layers of entangling gates (e.g., depth- k experiments, $b = x_{10}^S + k x_{ZI}^{\text{CNOT}} + x_{10}^M$) similar to cycle error reconstruction [17,27,28]. More details about the experimental construction are presented in Appendix B 2.

In practice, the experiments specified by F are each run many times to obtain an estimate for the vector $\mathbf{b}$. Multiple

rows $\{j\}$ of F may be estimated from one experimental setting, provided the Paulis $\{o_j\}$ are site-wise commuting and the circuits are the same. Minimizing the residual error in the least-squares problem $\|F\mathbf{x}_0 - \mathbf{b}\| \leq \epsilon$, where ϵ is chosen based on the tolerable amount of residual errors, produces a solution $\mathbf{x}_0$. The final estimate yields $\mathbf{x}_\eta = \mathbf{x}_0 + \mathbf{y}_\eta$, where $\mathbf{y}_\eta$ is a gauge vector depending on the gauge parameters η , associated with the kernel of F .

The key difference between this approach and previous attempts at Pauli noise learning [9,17,27–29] is that we consider the full gate set, as opposed to only subsets of it. Though each Pauli noise channel can only be determined up to a gauge transformation, they are related by the same gauge parameters η . Our approach resembles a technique known as averaged circuit eigenvalue sampling (or ACES) [30–33] which solves a system of linear equations similar to Eq. (9). A key difference is, whereas ACES constructs a full-rank design matrix by introducing additional assumptions and may not learn every learnable parameter, our design matrix fully characterizes all learnable parameters, leaving only the gauge undetermined.

B. Quasi-local noise models

The above framework can apply to larger systems by imposing a quasi-local noise model, such that the Pauli noise channels are determined by a number of parameters linear in the number of qubits. Such underlying assumptions about the locality of the noise are supported by experimental successes to date [9,10,21,28,34–37]. A quasi-local Pauli noise channel is defined as

$$\Lambda(\cdot) = \bigcirc_{a \in \mathcal{K}} (\omega_a P_a(\cdot) P_a + (1 - \omega_a)(\cdot)), \quad (10)$$

where $\bigcirc$ denotes composition of maps, and $\mathcal{K}$ is a set of local Pauli operators; that is, they are supported on a local subset of qubits. The ordering in this decomposition does not matter because the Pauli channels commute. Equivalently, we can define the generator of the channel $\mathcal{L}$ such that $\Lambda = e^{\mathcal{L}}$, where

$$\mathcal{L}(\rho) = \sum_{a \in \mathcal{K}} \frac{\tau_a}{2} (P_a \rho P_a - \rho). \quad (11)$$

We require $\omega_a < 1/2$ and define $\tau_a = -\log(1 - 2\omega_a)$ as the generator rate of the channel. Note that we allow τ_a (and thus ω_a) to be negative. We can then map the generator rates to log-fidelities through

$$x_a = \sum_{b \in \mathcal{K}} \langle a, b \rangle \tau_b, \quad \forall a \in \mathcal{P}^n, a \neq 0, \quad (12)$$

where $\langle a, b \rangle = 0$ if P_a commutes with P_b and 1 otherwise. For this relationship to hold, the set of local operators $\mathcal{K}$ should satisfy certain mathematical properties as explained

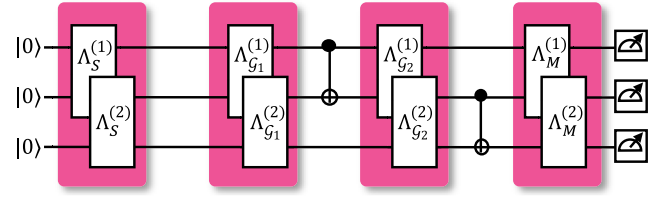


FIG. 3. Quasi-local noise model. An example of a 3-qubit system with 2-local noise generators.

in Appendix B 1. Another equivalent understanding is that a quasi-local Pauli channel can be expressed as a composition of (possibly non-positive) Pauli channels supported on local subsystems, as shown in Fig. 3. The noise parameters are given by τ , a vector comprised of all generators $\{\tau_a\}$ from each noise channel. Equivalently, the noise parameters can be chosen as $\mathbf{x}$ containing only x_a such that $a \in \mathcal{K}$, which is related to τ by an invertible matrix as in Eq. (12). Another useful way to parameterize a quasi-local Pauli channel is through the Möbius inversion of $\mathbf{x}$ [21,36], denoted by $\mathbf{r}$, which is a vector of length $|\mathcal{K}|$ interchangeable with τ by an invertible matrix. We define $\mathbf{r}$ in Appendix B 1. This definition of quasi-local Pauli models is widely used in the literature [9,21,36] under the name of sparse Pauli-Lindblad models (11) or inclusive Pauli channels, while there also exist alternative inequivalent definitions [28,37].

One major advantage of our definition is that the learnable and gauge parameters can be exactly characterized as a linear space over the noise parameters τ due to the linear relation between τ and $\mathbf{x}$ [21]. Specifically, as proven in Ref. [21], for all the quasi-local Pauli noise models considered in this work, the gauge parameters can be completely described by n single-qubit depolarizing channels, leading to a reduction in the number of gauge parameters from $2^n - 1$ to n . To learn such a quasi-local Pauli noise model up to gauge parameters, we will similarly construct a linear system of equations $\mathbf{b} = F\mathbf{x}$ (or $\mathbf{b} = F''\tau$) such that the design matrix F (or F'') reaches the maximal rank determined by the number of gauge parameters (See Appendix B 2 for more details).

III. THEORY RESULTS

A. Self-consistent error mitigation

A major motivation for learning quantum noise is to improve the performance of quantum computations. Quantum error mitigation (QEM) is one approach for reducing the bias in noisy quantum computations by utilizing information about the learned noise. As we discussed above, there exists a fundamental ambiguity in quantum noise learning on account of gauge degrees of freedom, leading to previous challenges for error mitigation [21]. Here, we will discuss how this limitation impacted existing QEM protocols, and how we overcome

the challenge by introducing a self-consistent QEM framework.

In this work, we focus on probabilistic error cancellation (PEC), one of the few QEM protocols with provable guarantees for achieving bias-free estimates of expectation values given sufficient samples and accurate noise characterization [5,11,38]. In particular, we study PEC protocols based on Pauli noise models [9,24]. Because PEC is provably unbiased under an accurate noise model, it provides the cleanest setting to isolate and quantify the effect of noise model inconsistencies. However, our self-consistent noise learning approach is not specific to PEC and applies to any error mitigation method that relies on a learned noise model, such as zero-noise extrapolation (ZNE) [3,39] or tensor-network error mitigation (TEM) [10,40]. The extension to ZNE is discussed in Appendix A 1.

Let us briefly review how PEC works: consider the task of expectation value estimation for an observable on the output state of a quantum circuit, which is a natural task in, e.g., Hamiltonian simulation. If one runs the circuit on real quantum hardware, noise would corrupt the expectation value. To retrieve the noiseless value, a naïve approach would be to cancel out all the noise channels Λ by implementing their inverse map Λ^{-1} . The challenge is that Λ^{-1} is generally not completely positive, thus not a physically realizable quantum channel. Nevertheless, when Λ is a Pauli channel, its inverse can be formally written as $\Lambda^{-1}(\cdot) = \sum_a p_a^* P_a(\cdot) P_a$ with $\sum_a p_a^* = 1$ but p_a^* can be negative. To implement Λ^{-1} in expectation, one can sample and apply a Pauli gate P_a according to the following distribution,

$$q_a = \frac{|p_a^*|}{\gamma}, \quad \text{where } \gamma = \sum_a |p_a^*|, \quad (13)$$

where a factor of $\gamma \cdot \text{sign}(p_a^*)$ is then multiplied with the experimentally measured estimator, resulting in the cancellation of the noise channel Λ in the expectation value. Applying this procedure to cancel every noise channel in the circuit, the resulting estimation is an unbiased estimator for the noiseless expectation value. While this procedure works for *any* quantum circuit, the trade-off is that an additional sampling overhead of $\prod_j \gamma_j^2$ is needed where γ_j corresponds to the j th noise channel, due to the multiplied pre-factors.

Computing p^* can be computationally challenging in general. For quasi-local Pauli channels as in Eq. (10), an alternative approach is given in Ref. [9]. First note that the inverse of Λ can be written as

$$\Lambda^{-1} = \bigcirc_{a \in \mathcal{K}} \left(\frac{-\omega_a}{1-2\omega_a} P_a(\cdot) P_a + \frac{1-\omega_a}{1-2\omega_a} (\cdot) \right). \quad (14)$$

Then, we can simply invert each factor of Λ^{-1} . For $\omega_a \leq 0$, the factor is a proper Pauli channel that can be implemented without any overhead. For $0 < \omega_a < 1/2$, the

factor can be implemented in expectation with an overhead $\gamma_a = 1/(1-2\omega_a) = \exp(\tau_a)$. Multiplying the overhead from each factor yields,

$$\gamma = \exp \left(\sum_{a \in \mathcal{K}} \max(0, \tau_a) \right), \quad (15)$$

which we refer to as the overhead associated with the quasi-local Pauli channel Λ .

The above PEC protocol requires full knowledge of the noisy gate set so as to implement the inverse noise channels. However, there generically exists gauge ambiguity in learning the noise parameters, hindering a direct application of PEC. In recent literature of error mitigation with Pauli noise models [3,9,24], the issue of gauge ambiguity is circumvented by imposing the “symmetry assumption.” As an illustration of this assumption, we consider a Clifford gate $\mathcal{G}$ that satisfies $\mathcal{G}^2 = I$ where the gate can be, for example, a CNOT gate. For any $P \in \mathcal{P}^n$, $Q = \mathcal{G}(P)$, whenever $P \neq Q$ up to a sign, the symmetry assumption imposes $\lambda_P^{\mathcal{G}} = \lambda_Q^{\mathcal{G}}$ (e.g. $\lambda_{XI}^{\text{CNOT}} = \lambda_{XX}^{\text{CNOT}}$). This ensures every $\lambda_P^{\mathcal{G}}$ can be uniquely determined by cycle benchmarking [17,27,28]. Furthermore, the state-preparation noise is assumed to be noiseless to determine and mitigate measurement noise [19,41]. We later show that these assumptions are not only unnecessary, but also lead to *inconsistent* characterization of the gate set, which results in biased estimates of expectation values in applications such as QEM.

Here, we propose a self-consistent PEC protocol that properly considers the gauge parameters, thus disambiguating Pauli noise in quantum computers. Our protocol builds on the gate set Pauli noise learning framework [21] discussed in the last section, which enables learning a set of noise channels $\Lambda_\eta = \{\Lambda_\eta^S, \Lambda_\eta^M, \{\Lambda_\eta^{\mathcal{G}}\}\}$ that are gauge-equivalent to the true noise channels $\Lambda = \{\Lambda^S, \Lambda^M, \{\Lambda^{\mathcal{G}}\}\}$, meaning that the two noisy gate sets have exactly the same behavior in any experiments. Thus, without ever fixing the gauge parameters, the learned Λ_η contains as much information as the true noisy gate set, which can be applied to PEC. We formalize our claim in the following theorem.

Theorem 1. (Self-consistent PEC). Let $\Lambda_\eta = \{\Lambda_\eta^S, \Lambda_\eta^M, \{\Lambda_\eta^{\mathcal{G}}\}\}$ be a collection of Pauli noise channels that are gauge-equivalent to the true noise channels. By applying PEC as if Λ_η is the ground truth, one can obtain unbiased estimators for any circuits and observables.

The formal statement and proof for Theorem 1 is given in Appendix A. An illustrative proof is provided for a two-qubit system in Fig. 4. Note that Theorem 1 holds even for quasi-local Pauli noise models, which is crucial for scaling up to large systems.

In light of Theorem 1, it is straightforward to explain how the assumptions of symmetric gates and perfect initialization lead to inconsistency. Basically, the

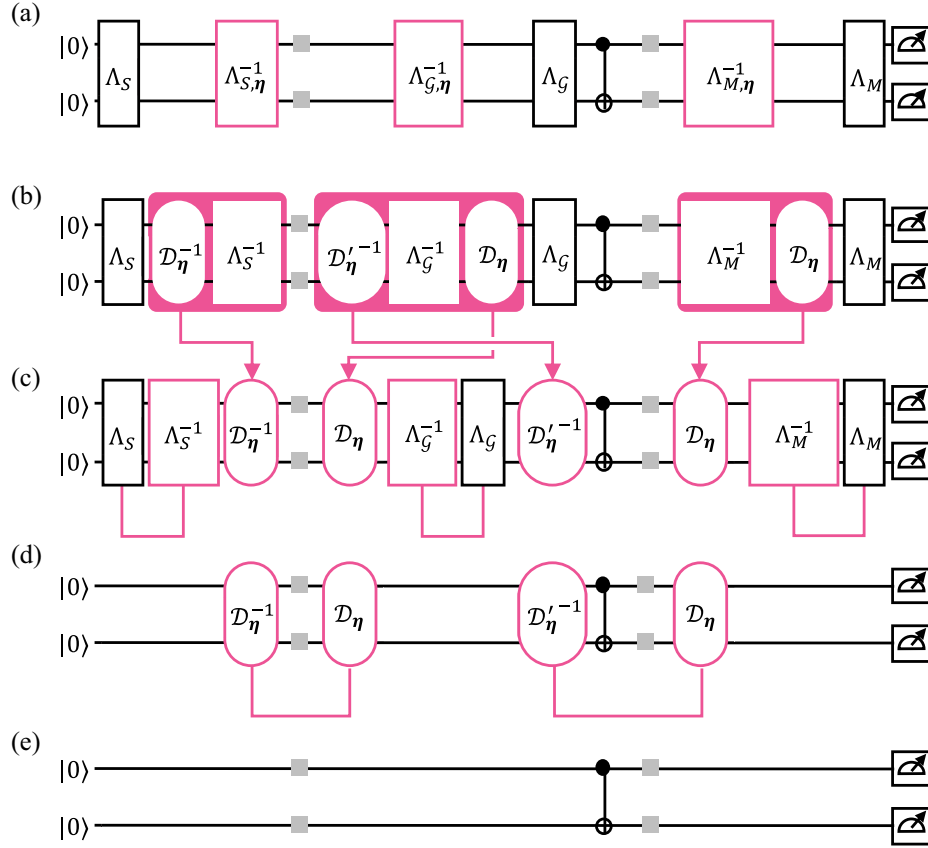


FIG. 4. Graphical proof that a complete gate set learned in a self-consistent manner can be validated using an error mitigation formalism. (a) Using the probabilistic error cancellation (PEC) framework, the quasi-inverse channels $\Lambda_{S/G/M,\eta}^{-1}$ can be applied at the expense of a sampling overhead which increases exponentially with the amount of noise in the constituent noise channels. Unlike previous formulations, we apply this inverse channel in a self-consistent manner where the entire gate set shares the same set of gauge parameters η . For the sake of clarity, we chose a controlled-NOT gate as the two-qubit gate. This same proof applies to any other type of two-qubit gate, and to arbitrary numbers of qubits. (b) Substitution of the gauge $\mathcal{D}_\eta$ and Pauli noise channels $\Lambda_{S/G/M}^{-1}$ from Fig. 1(b) yields the operations in the red boxes. (c) Reordering the gauge channels (pink arrows) leads to cancellations, or compositions of identity channels (pink brackets). Note that Pauli channels commute with each other. (d) The gauge channels and their inverses also compose to identity channels (pink brackets). Note that the generalized depolarizing channels commute with any single-qubit gates (gray squares). Also recall $\mathcal{D}'_\eta$ is defined to be the gauge channel conjugated by the entangling Clifford gate ($\mathcal{G} \circ \mathcal{D}'_\eta = \mathcal{D}_\eta \circ \mathcal{G}$). (e) Finally, the resulting, mitigated circuit shows a noise-free operation of a controlled-NOT gate on two qubits.

assumptions result in fixing each Pauli channel individually, leading to a model in the form of $\{\Lambda_{\eta_S}^S, \Lambda_{\eta_M}^M, \{\Lambda_{\eta_G}^G\}\}$, where an inconsistent choice of gauge parameters— η_S, η_M, η_G —is applied for each component of the gate set. To see this in the context of Fig. 4, the residual generalized depolarizing channels remaining after mitigation would result in a biased estimation.

We remark that the idea of combining gate set tomography (GST) [13,42] with QEM to address gauge ambiguity has been discussed in the literature [12]. However, due to the extreme complexity and resource cost of GST, it is unclear how to apply such protocols beyond a few qubits. Instead, our method builds on the recently proposed gate set Pauli noise learning framework [20,21], which enables explicit and efficient parameterization of all learnable and gauge parameters under a practical quasi-local noise assumption. To our knowledge, this is the first

experimental demonstration of self-consistent QEM, with comparable scalability as state-of-the-art QEM protocols [3,9,24].

Finally, recall PEC requires a sampling overhead $\prod_j \gamma_j^2$. Theorem 1 suggests that, by assuming the true noise model to be any of the gauge-equivalent models Λ_η , parameterized by the gauge parameters η , PEC yields unbiased estimators. Interestingly, while different Λ_η all yield the same observable outcomes, different gauge parameters do not give us the same sampling overhead. This motivates us to conduct gauge optimization—searching for $\hat{\eta}^*$ that minimizes the PEC overhead. The reader may wonder whether the difference in sampling overhead can be used to determine the gauge parameters η . This is not possible, as the overhead merely depends on what we infer the noise parameters to be, but not what the true parameters actually

are. In this work, we unify the gate set learning with shot noise and gauge optimization as a single convex optimization problem, which is efficiently solvable and can drastically reduce the PEC sampling overhead.

We provide an experimental demonstration of mitigation overhead reduction through gauge optimization in Sec. IV D.

B. Efficient gauge optimization

Beyond addressing potential sources of bias in mitigated expectation values, we now show how the self-consistent learning approach can also be used to reduce the sampling complexity needed to successfully perform error mitigation for *any* quantum circuit. Recall that we have $\mathbf{b} = F\mathbf{x}$; however, in this case we prefer the noise parameters in the basis of the reduced parameters $\mathbf{r}$ because it is polynomial in size for quasi-local models. The conversion from $\mathbf{x}$ to $\mathbf{r}$ is discussed in Appendix B 1, which allows us to rewrite the design matrix as $\mathbf{b} = F'\mathbf{r}$. Suppose we have a design matrix F' and estimation of $\hat{\mathbf{b}}$ from experiments. A naïve approach to obtaining $\mathbf{r}$ and minimizing the sampling overhead needed for QEM involves first performing a pseudo-inverse of the design matrix $\hat{\mathbf{r}}_0 = F'^+\hat{\mathbf{b}}$ (which fixes the residual error $\epsilon = \|F'\hat{\mathbf{r}}_0 - \hat{\mathbf{b}}\|_2$), followed directly by a second optimization step over gauge parameters $\boldsymbol{\eta}$ on the overhead $\gamma = \exp(\sum_{a, \text{layer}} \max(\tau_a^{\text{layer}}, 0))$ where the sum is performed over all quasi-local generators τ_a for all layers of gates, see Eq. (15). However, such an approach unnecessarily restricts the gauge optimization procedure without taking into consideration that the residual errors can vary depending on the statistical fluctuations of the observed outcomes.

Rather, we introduce a *one-step* optimization strategy where the possible $\mathbf{r}$ parameters are searched in a self-consistent manner subject to a constrained residual error ϵ chosen *a priori*. That is, we solve the following optimization problem:

$$\min_{\hat{\mathbf{r}}} \left\{ \sum_{a \in \mathcal{K}, \text{layer}} \max(\hat{\tau}_a^{\text{layer}}, 0) \right\} \text{ s.t. } \|F'\hat{\mathbf{r}} - \hat{\mathbf{b}}\|_2 \leq \epsilon. \quad (16)$$

where the parameters $\hat{\tau}_a$ depend on $\hat{\mathbf{r}}$ following Eqs. (12) and (B4). In Appendix D 3, we show the above gauge optimization problem can be rewritten as a second-order cone program (SOCP), thus provably solvable in polynomial time with n using standard convex optimization solvers [43,44], as long as the quasi-local model we use contains only polynomially many parameters.

IV. EXPERIMENTAL RESULTS

In the following we report a series of experiments that demonstrate the importance of self-consistent noise learning for error mitigation with increasing complexity of the

noise models. In all of these experiments, we learned two noise models. The first model represents the previous state-of-the-art [3,9,19] which imposes the symmetry assumption between conjugate Pauli eigenvalues and assumes ideal state preparation for readout error mitigation. We refer to this as the “inconsistent” noise model. The second model learns all noise channels in a self-consistent way and is referred to as the “consistent” noise model. We note that in Secs. IV A–IV C, we do not physically implement PEC. Instead, we compare the performance of both models in predicting noisy expectation values in the corresponding circuits, which is numerically tractable due to the Clifford nature of the circuits. The ratio of the predicted and measured estimates the bias that a hypothetical PEC implementation would exhibit. An end-to-end experimental demonstration of PEC, including actual quasi-probabilistic circuit sampling, is presented in Sec. IV D.

A. Restricted experiment on two qubits

As a first step, we examined the impact of self-consistent learning for a gate set on two qubits. Recall that methods which use a framework lacking this consistency may work on some experiments but fail on others. To highlight the difference between such learning frameworks, we used a 27-qubit device named *ibm_auckland* with calibrated CNOT gates as the basis two-qubit gates. In this case, our full gate set was composed of initialization to $|0\rangle^{\otimes 2}$, a single CNOT gate, any single-qubit gates (with negligible noise), and computational basis measurements. A rigorous way to represent all possible experimental outcomes on this two-qubit system is captured in the pattern transfer graph described earlier (Fig. 2).

Rather than examining all the cycles in the pattern transfer graph, we prepared and measured experiments in the Z-basis exclusively, and thus we needed to focus on learning only two specific cycles: the $ZI\bar{I}$ and $IZ \leftrightarrow ZZ$. While one of these cycles involves only a single node (“10”), the other involves two nodes (“01” and “11”). The latter is referred to as a degenerate cycle or a conjugate pair as they contain two eigenvalues that cannot be separately determined [9,20,24].

A convenient consequence for focusing on the Z-only eigenvalues means the preparation and measurement bases can be entirely in the computational states (i.e. $|0\rangle$ or $|1\rangle$). Thus, to learn cycles restricted to a certain type—in this case the Z-only observables—we only needed to prepare the $|00\rangle$ initial state for circuits with increasing repetitions of the CNOT gate from depths 0 to 32. Unlike previous noise learning approaches which only utilized circuits with even numbers of CNOT gates, here we also introduced a learning circuit with just a single, or depth-1, application of CNOT. For larger experiments discussed later, more preparation and measurement bases for the depth-1 experiments need to be incorporated to learn all possible

learnable parameters described under Eq. (9). Intuitively, the reason depth-1 experiments are needed for self-consistent learning is to account for the degeneracy between the conjugate Paulis—in this case between IZ and ZZ . Thus for the same set of experiments, we were able to learn two noise models for the eigenvalues λ_{ZI} , λ_{IZ} , and λ_{ZZ} : one that is *self-consistent* and another *inconsistent* which assumes that any conjugate Paulis are symmetric and does not incorporate depth-1 results.

To compare the validity of the two noise learning approaches, we then separately performed so-called “target” experiments where the observables $\langle ZI \rangle$ and $\langle ZZ \rangle$ were measured after initialization into the $|11\rangle$ state; we compared the outcomes against those predicted from the

learned noise models using only the $|00\rangle$ initial state. Ideally, the experimentally measured outcomes should agree with those predicted from the noise models, and thus their division should yield an ideal value of 1—known in this case because these are Clifford operations on an initial stabilizer state. The ratio of the two values—measured to the predicted—informs us how much mitigation bias persists based on the different noise models.

We show the experimental outcomes, along with those predicted by the two different noise learning procedures, in Figs. 5(b), 5(c). As expected, the non-degenerate cycle involving the $\langle ZI \rangle$ observable exhibited no difference in bias between the experimentally measured and the predicted outcomes from two noise learning approaches at

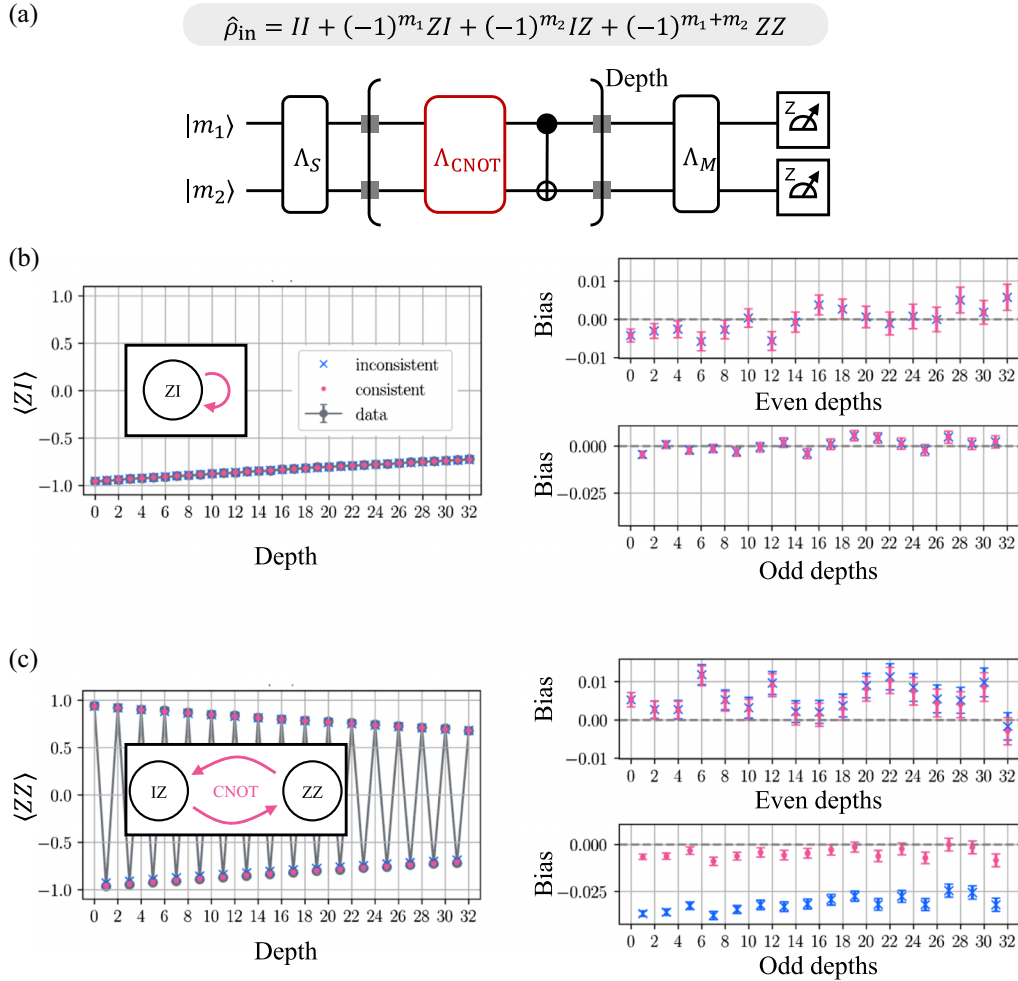


FIG. 5. Experimental learning and mitigation of a restricted set of errors on two qubits. (a) Circuit used for both learning and targeted mitigation, except an initial state with $m_1 = m_2 = 0$ is used for the learning while $m_1 = m_2 = 1$ is used for the target circuit. This restricts learning to three Pauli terms: IZ , ZZ and ZI . Note that the target circuit is prepared in $|11\rangle$, the $-IZ$ and $+ZZ$ eigenstates, which differs from the learning circuit which is prepared in $|00\rangle$, the $+IZ$ and $+ZZ$ eigenstates. The “inconsistent” noise model only requires learning with circuits using even depths (0, 2, 4, ..., 32), while the “consistent” learning model requires one additional depth-1 experiment. (b, c) For the non-degenerate (b) and degenerate (c) cycles, the experimental data (gray) from the target circuit are shown alongside the predicted outcomes using the inconsistent (blue) and consistent (pink) noise models. To the right of (b), both the even and odd depths show no difference in predicted outcomes. However, to the right of (c), the even depth shows no difference while the odd depths show a difference of $3.2 \pm 0.4\%$ (blue, inconsistent) compared to a $0.5 \pm 0.3\%$ (pink, consistent) bias in the outcomes.

even or odd depths. The reason we separated out the even from odd depths is because the “inconsistent” noise model, in this special case, is unambiguous at predicting outcomes of circuits for depth-even applications of CNOTs.

In contrast, the degenerate cycle $IZ \leftrightarrow ZZ$ showed no mitigation bias when an even number of CNOT gates were applied, as expected based on the fact that the “inconsistent” model accurately captures the product of the conjugate Pauli eigenvalues. However, for odd-depths applications of the CNOT gate, we find that the self-consistent learning protocol reduces the predicted bias from $3.2 \pm 0.4\%$ down to $0.5 \pm 0.3\%$ when averaged over all 16, odd layer depths of the target circuit up to depth 31. This statistically significant improvement in predicted outcomes using the self-consistent learning approach was reproduced across a total of six qubit pairs on the same device (Data in Appendix C 1). Despite the restricted nature of this experiment on only two qubits, the widespread improvement in mitigation bias with self-consistent learning motivated the question of how much this bias can be improved for circuits with more qubits.

B. Restricted experiment on entangled states

Next, we examined the impact of self-consistent noise learning for a target circuit with not only many more qubits, but also an observable that depends mostly on individual fidelities from degenerate cycles. We identified an observable such that the observed bias should increase with system size when compared against the “inconsistent” noise learning approach. Meanwhile we expect the “consistent” noise learning approach, which captures the degenerate Pauli pairs, to remain unbiased no matter the system size.

For this task, we identified the highly entangled GHZ state as the ideal target state. The GHZ state on n qubits is a stabilizer state that is specified by being the simultaneous $+1$ eigenstate of the generators, a set which includes a full-weight term $\langle X^{\otimes n} \rangle$ and $n - 1$ weight-2 terms $\langle Z_i Z_{i+1} \rangle$ for $i \in [n - 1]$. However, rather than relying on well-known preparation circuits which require learning $O(n)$ unique layers of entangling gates, we prepared the state using only two unique dense layers of simultaneously applied CNOT gates. Specifically, we first chose a fixed set of 21 qubits. We then employed a SAT-solver to prepare GHZ states on $n = 1, 3, \dots, 21$ -qubit subsets of those 21 qubits using the two unique layers of entangling gates covering all 21 qubits [45,46]. A graphical way to check how our procedure works can be seen in Figs. 6(a) and 6(b) for $n = 5$ and $n = 21$ GHZ states, respectively. In those figures, we track the creation of the $\langle X^{\otimes n} \rangle$ stabilizer and show that it grows monotonically with densely populated CNOT layers of gates.

We call the two alternating template layers “a” and “b.” We fix the CNOT directions for each template layer and use interleaving single-qubit gates to effectively change

the CNOT directions to arrive at the circuits in Fig. 6. For this experiment where we are only examining the impact of self-consistent learning on the $\langle X^{\otimes n} \rangle$ observable of the target GHZ state, we only learned the Pauli eigenvalues which contribute to the construction of the final observable. This was only possible because our target circuit is a Clifford circuit, which means we were able to classically back-propagate the observable through the entire circuit and identify those Pauli eigenvalues needed from each instance of the two template layers. In this sense, this was a *restricted* noise model because we did not learn the full Pauli noise channel for both layers, but allowed for the possibility of *non-local* noise by not imposing any locality constraints; in other words, we allowed the number of gauge parameters to remain in the most general form with $2^n - 1$ terms. In Appendix B 2, we include an example for learning the noise of template “a” used in preparing the $n = 21$ GHZ state.

Unlike the previous section, the experiments here and in the subsequent sections were conducted using a larger, 127-qubit device named *ibm_strasbourg*. Similar to how we compared the learned noise models against the target circuit earlier, we again compare the predicted outcomes for the $\langle X^{\otimes n} \rangle$ observable for GHZ states with increasing sizes up to 21 qubits against the experimentally measured values [Figs. 6(c), 6(d)]. The Clifford nature of the circuit allows us to predict the resulting bias of a hypothetical mitigation experiment with PEC by dividing noisy expectation values by the values predicted by the respective models. We refer to these as “mitigated” values. Indeed, for GHZ states up to 21 qubits, we observed an increasing bias using the inconsistent noise model reaching $35.2\% \pm 6.5\%$, while we observed statistically insignificant $-1.2\% \pm 4.1\%$ biases using the self-consistent noise model for the largest depths.

C. Scalable learning for general, quasi-local noise

Earlier we focused on restricted models with learning circuits that are straightforward to construct; now we will focus on complex learning circuits with minimal assumptions needed for constructing the full noise models. That is, the previous two experiments used some knowledge of the target circuit to inform the design of the learning experiments, while in this section we will discuss how to conduct complete self-consistent noise learning when given only the quasi-local noise assumption based on qubit connectivity and the template gate layers being learned. By providing an explicit construction for the gate layers in the gate set and a noise ansatz (e.g. 1- or 2-local), we used the formalism shown in Fig. 3 and in Ref. [21] to construct the preparation and measurement bases needed for the learning circuits such that a self-consistent noise model could be inferred. Since the noise is assumed to be quasi-local, the number of parameters is no longer exponential but in fact only linear, and thus can be efficiently learned.

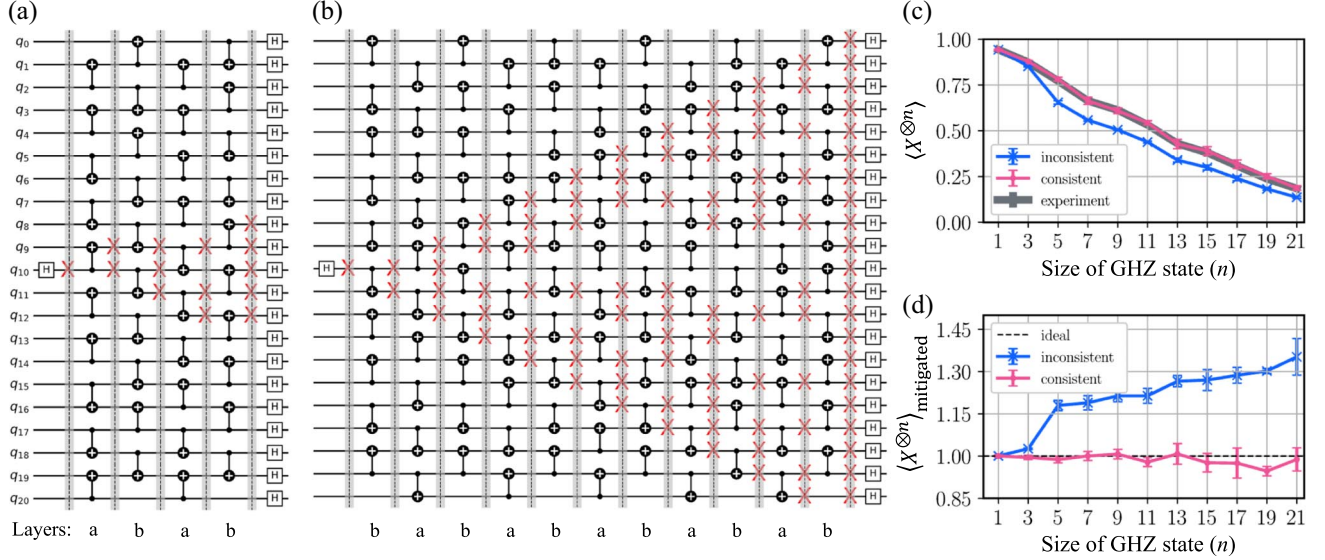


FIG. 6. Self-consistent experimental learning of non-local errors and mitigation of a GHZ state on up to 21 qubits. (a) Quantum circuit used to prepare a 5-qubit GHZ state using two template layers: “a” and “b,” where layers from the same template do not necessarily have the same direction of controlled-NOT gates. The root qubit, where the a Hadamard (H) gate is applied before the first layer, can be described by a weight-1 stabilizer X (red), which is shown to grow to weight-5 after four total layers. (b) Quantum circuit used to prepare a 21-qubit GHZ state using the same two template layers, except the controlled-NOT gate directions again do not match with the circuit used to prepare the 5-qubit GHZ state in (a). The growth of the $\langle X^{\otimes n} \rangle$ stabilizer for the GHZ state reaches full-weight by the end of the circuit as indicated by the column of red X labels immediately before the final layer of Hadamard gates. (c) Experimental outcomes (black line) plotted against predicted expectation values using self-consistent (pink circles) and inconsistent (blue crosses) noise models. (d) Mitigated values of the full-weight, $\langle X^{\otimes n} \rangle$ stabilizers of the GHZ state for circuits up to $n = 21$ where the self-consistent noise model (pink circles) shows strong agreement with the expected value of $\langle X^{\otimes n} \rangle = 1$ while the symmetric noise model (blue crosses) shows increasing bias with the size of the GHZ state. The error bars for both (c) and (d) were the result of averaging over seven separate experiments (Appendix C 2 for more details).

For n qubits on a ring, we consider a gate set consisting of two gate layers $G_{\text{even}} = \text{CNOT}_{1,2} \otimes \text{CNOT}_{3,4} \otimes \cdots \otimes \text{CNOT}_{n-1,n}$ and $G_{\text{odd}} = \text{CNOT}_{2,3} \otimes \text{CNOT}_{4,5} \otimes \cdots \otimes \text{CNOT}_{n,n+1}$. The noise on each layer and on SPAM operations is assumed to be *quasi-local*, i.e., it factors into a composition of channels that act only on nearest neighbor qubits. As shown in Ref. [21], this model has $28n$ parameters with a fully local gauge. That is, there are $27n$ learnable parameters and n gauge parameters corresponding to n single-qubit depolarization maps.

Due to the locality of the noise model, only local expectation values are needed to learn the model parameters. This allows many parameters to be estimated in parallel. As a result, the number of measurement settings needed to learn the complete noise model remains constant and does not scale with system size. For details of the experiment and specific measurement settings, see Appendix C 3 and Table II.

We applied the learned noise model to a target circuit where we measure local Z observables for every qubit on a closed ring. As before, the circuit consists of two layers of CNOT gates between even or odd neighboring qubit pairs. Specifically, this is a Clifford circuit designed in a way such that the 92 observables each depend only on

weight-1 and weight-2 Pauli eigenvalues which all originate from different degenerate cycles and are thus sensitive to the symmetry assumptions imposed by the inconsistent model. In total, the Pauli eigenvalues probed by the observables cover all degenerate Pauli eigenvalues of the participating CNOTs (See Appendix C 3 for details). We used a ring of 92 of the 127 qubits available on *ibm_strasbourg* shown in Fig. 7(a). With every layer of two-qubit blocks, each weight-1 eigenstate is propagated to another weight-1 eigenstate shifted by one qubit index along the ring. Note that one such block consists of two layers of parallel CNOT gates as depicted in Fig. 13(b). We highlight one of the 92 available observables and show how it evolved for different circuit depths in Fig. 7(b). Then, we compare the experimental outcomes against the predicted outcomes based on the consistent and inconsistent noise models by computing the mitigated values as before. In Fig. 7(c), we show one specific example where the bias using the inconsistent model reaches $12\% \pm 0.5\%$ whereas the consistent model shows no statistically significant bias of $0.3 \pm 0.5\%$. Applying the same analysis as described above across all 92 qubits, we saw that the consistent noise model yielded mitigation errors at or below that predicted with the inconsistent noise model

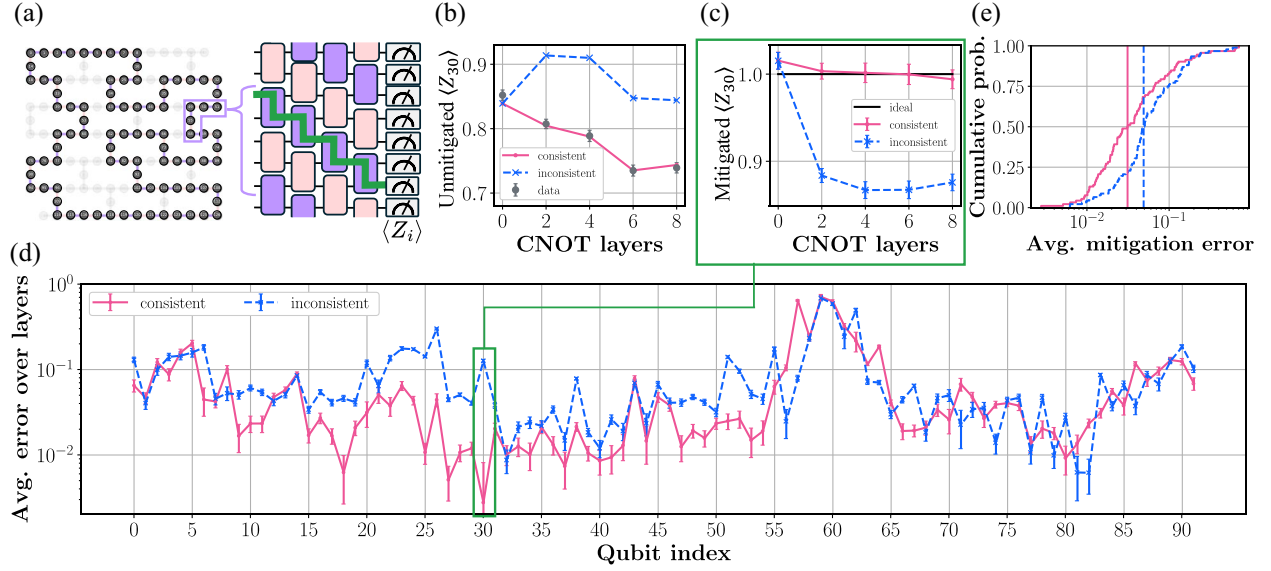


FIG. 7. Scalable (quasi-local) self-consistent noise learning and mitigation of weight-1 observables on a ring of 92 qubits. (a) Closed loop of 92 qubits on a 127-qubit device, *ibm_strasbourg*. Boxed section of ring shows a cross section of the 92 qubit, depth-8 quantum circuit designed specifically to propagate all 92, weight-1 stabilizers in a “staircase” (green) fashion such that the initial and final qubit support of the stabilizer falls on a different qubit. (b) Experimentally measured (filled gray circle) expectation values versus number of circuit layers compared against the self-consistent (solid red) and symmetric noise predictions (dashed blue). (c) Predicted values divided by the measured values yield a mitigated value for the data set in (b), and boxed in (a). (d) Average mitigation error up to four circuit layers on all 92 qubits calculated in the same manner as for a single qubit as seen in (c). Error bars depict one standard deviation of the shot noise on the unmitigated data. (e) Cumulative distribution of mitigation errors (d) between the experimentally measured and predicted outcomes. The median mitigation error, denoted by vertical lines, shows a reduction from 4.9% (dashed blue) to 3.1% (solid red) bias.

[Figs. 7(d), 7(e)]. In fact, the median mitigation error was reduced from 4.9% to 3.1%. The remaining bias can be largely attributed to out-of-model errors [47]. More specifically, there are several physically relevant error sources that fall outside the two-local Pauli-Lindblad noise ansatz used here. Non-Markovian effects, such as coupling to two-level systems (TLS) [48], slow temporal drift of the noise parameters between learning and target circuits, or coherent leakage out of the qubit subspace, are not captured by our noise model. Similarly, while we allow arbitrary two-body interactions between nearest-neighbor qubits, three-body correlations or long-range coherent crosstalk between non-nearest-neighbor qubits lie outside the quasi-local ansatz. While some of these errors, such as long-range crosstalk, can in principle be twirled into Pauli channels, the resulting channels are not guaranteed to be representable by a quasi-local Pauli-Lindblad generator. Others, such as leakage, cannot be twirled into Pauli noise at all. In either case, such errors fall outside the model class assumed here. As analyzed in Ref. [47], the resulting mitigation bias from non-local interactions scales linearly with the number N of such terms as $\sim 4N\lambda$, where λ characterizes their strength, implying that these effects can become more pronounced at larger system sizes. In practice, temporal drifts can be mitigated by more frequent noise characterization, while non-local crosstalk can be

substantially suppressed using context-aware dynamical decoupling and compilation techniques [49]. We stress that these limitations are shared by all model-based error mitigation approaches, including PEC, ZNE, and TEM, and are not specific to the self-consistent framework introduced here. Rather, the present work eliminates gauge inconsistency as a dominant and previously unaccounted-for source of systematic error, so that the remaining bias can be more cleanly attributed to these genuine physical model violations rather than to artifacts of the learning and mitigation pipeline.

D. Gauged-optimized PEC

The experiments presented so far have established that inconsistently learned noise models lead to systematic biases when predicting noisy expectation values. In the previous sections, this effect could be cleanly identified because the employed benchmark circuits were Clifford, allowing for exact classical predictions of the noisy outcomes. We now go one step further and demonstrate that the same inconsistencies lead to biased *mitigated* expectation values when applying PEC (see Sec. III A) to non-Clifford circuits. Moreover, our experiment shows that gauge optimization can substantially improve the sampling overhead of PEC.

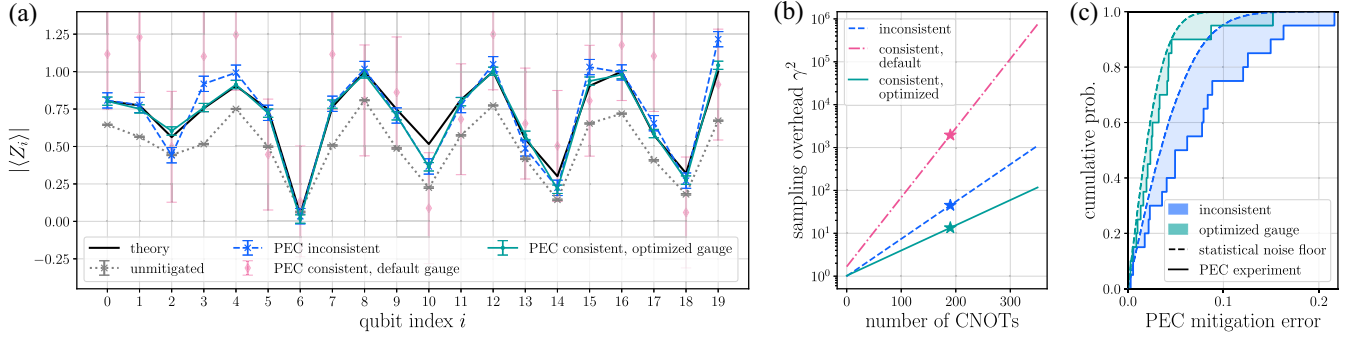


FIG. 8. Probabilistic error cancellation with self-consistent noise models and gauge optimization. (a) PEC results for single-qubit Z observables in a 20-qubit non-Clifford benchmark circuit with 20 CNOT layers. Error bars show the sample standard deviation of the PEC estimators. (b) The PEC sampling overhead γ^2 as a function of the number of CNOT gates for the characterized noise models, where the star corresponds to the circuit from panel a. (c) Cumulative distribution of the absolute remaining error of the PEC estimators for the inconsistent noise model and the gauge-optimized noise model (solid lines). Dashed lines show the statistical noise floor of unbiased PEC from repeated numerical simulations with the same number of shots and γ -values. Shaded regions reflect the systematic error beyond what is expected from shot noise alone.

As benchmark circuits, we considered the circuits introduced in Sec. IV C, but perturbed them away from Clifford angles as described in Appendix C 4 and again focused on single-qubit Z operators. All experiments were performed on the *ibm_fez* device at a system size of 20 qubits, for which exact classical predictions of the ideal expectation values remain tractable. The noise affecting these circuits was characterized using the same procedure as in Appendix D 2, assuming a quasi-local sparse Pauli-Lindblad noise model.

For the self-consistently learned noise model, we performed gauge optimization following the method described in Sec. III B. The fitting tolerance ϵ in Eq. (16) was chosen to be consistent with the shot noise present in the learning data, as detailed in Appendix D 2. For the benchmark circuits considered here, this results in markedly different PEC overheads: the inconsistent model yields a mitigation overhead of $\gamma = 6.74$, the self-consistent model obtained via direct inversion of Eq. (9) (which we refer to as the *default* gauge choice) yields $\gamma = 48.68$, while the gauge-optimized self-consistent model achieves a substantially reduced overhead of $\gamma = 3.7$. We emphasize that the dramatic overhead reduction compared to the default gauge reflects the suboptimality of that gauge within the self-consistent framework. The gauge-optimized model also achieves a reduction in overhead relative to the inconsistent model, but a direct comparison of γ values between them is not straightforward, since the latter produces biased estimators. The inconsistent model's lower overhead relative to the default consistent gauge comes at the cost of biased estimators, meaning its apparent efficiency reflects an unaccounted systematic error.

We applied PEC using these three noise models, see Fig. 8(a) for the resulting mitigated expectation values. For each model, PEC was performed using 100,000 randomized circuit instances, which include both

quasi-probabilistic sampling of the inverse noise channels and Pauli twirling, with a single shot taken per instance. While state preparation and gate errors are mitigated via genuine PEC sampling, measurement errors were corrected in post-processing in the spirit of TREX: We divided the estimators by the corresponding readout fidelities (or, in the case of the inconsistent model, by the inferred SPAM fidelities) associated with the measured observables. To further suppress non-Markovian effects, we employ a post-selection strategy that detects leakage out of the computational subspace [50,51]. After this procedure, we retained approximately 18 % of the 100,000 collected shots.

As expected, the mitigated values using the default self-consistent model exhibit very large statistical error bars due to its large associated γ value. The inconsistent model, on the other hand, displays a clear residual bias, which we attribute to the same systematic modeling errors identified in the previous sections. In contrast, the gauge-optimized self-consistent model performs best overall, simultaneously exhibiting lower bias and reduced variance. The difference in PEC sampling overhead for the three models is further illustrated in Fig. 8(b). We note the offset of the curves corresponding to the self-consistent models, which arises from the additional overhead associated with mitigating state preparation errors—an overhead that is absent by construction in the inconsistent model.

Quantitatively, PEC using the inconsistent model reduces the median error from 21.8% (unmitigated) down to 5.6%. This is further improved to a median error of 2.4% when using the gauge-optimized self-consistent model; the full error distributions are shown in Fig. 8(c). In the same figure, we compare the observed mitigation errors to the expected statistical noise floor obtained from simulations of unbiased PEC with the same number of samples and γ values. We find that the remaining errors for the gauge-optimized model are largely consistent with residual shot

noise, whereas the inconsistent model exhibits a pronounced deviation from this statistical noise floor, indicating the presence of additional systematic errors. This observation is consistent with the conclusions drawn from the Clifford experiments presented earlier in this work. Taken together, these results provide direct experimental evidence that gauge optimization can substantially enhance the practical performance of error mitigation, both by reducing bias and by significantly lowering the associated sampling overhead.

Our experiment represents one of the most rigorous realization of PEC to date, surpassing all previous implementations in the number of sampled circuit instances. Moreover, previous works on noise-model-based error mitigation have implored Markovian models with non-negative generator rates [3,9,10]. Our implementation generalizes this to non-Markovian noise models that include negative generator rates, as recently proposed in independent theoretical works [52,53].

V. DISCUSSION

As noise in quantum computers continuously improves, quantum error mitigation will become increasingly effective at unlocking some of the potential applications promised by fault-tolerant quantum computers [54]. While the idea of leveraging improved noise learning procedures for error mitigation was proposed [12], there was no proposal, to the best of our knowledge, to demonstrate the idea in a scalable manner. By showing that an accurate, self-consistent noise learning framework can be utilized for one of the most prominent error mitigation techniques, we have taken an important step toward realizing practically useful applications on pre-fault-tolerant quantum systems. While learning a quantum process by itself can be a candidate for quantum advantage [55], for example by utilizing entanglement to obtain a substantially lower sample complexity [56,57], it can also be used for more precise diagnosis of the most immediate hardware or material limitations to be addressed [58].

Our method resolves the issues arising from treating noise in different components of an experiment inconsistently, e.g., choosing different gauges for SPAM and gate errors. While it does not uncover the true, unobservable gauge, we anticipate that combining this approach with insights from the underlying physics of the processes can lead to a more accurate characterization of the actual noise affecting operations. For example, a more detailed understanding of entangling gates [59] and the different mechanisms involved in state preparation versus measurement may help determine a more physically meaningful gauge.

Our experimental approach, while requiring a constant number of additional circuits to learn the noise, results in verifiable accuracy for both deep and large circuits similar to those used in near-term and long-term application circuits. Unlike other approaches whose formalism

depends heavily on the locality of the physical noise, our experimental design can be catered to any noise ansatz based on the underlying quantum hardware. Furthermore, once the noise is accurately learned, an important application of this framework is that the associated sampling overhead needed for error mitigation can be reduced.

It would be intriguing to extend this self-consistent noise formalism to dynamic circuits where subsequent classical operations can depend on the outcomes of mid-circuit measurements [60,61]. Such dynamic circuits are considered promising for preparing and simulating interesting states with significantly less circuit depth [62–64]. For hybrid quantum-classical computations, dynamic circuits are also believed to be free of barren plateaus [65]. Being able to mitigate such mid-circuit measurement errors, once accurately learned in a self-consistent manner, can open up new avenues for quantum error mitigation [66,67]. More accurate noise models of such non-unitary operations are also essential for optimizing the performance of decoders needed to *actively* correct errors in large-scale, fault-tolerant quantum computers [4,68].

ACKNOWLEDGMENTS

We are grateful for helpful discussions with Zlatko Mineev, Maika Takita, Abhinav Kandala, Alexander Ivrii, David Layden, Ian Hincks, Sam Ferracin, James Raftery, Blake Johnson, Steve Flammia, Zhihan Zhang, Yunchao Liu, Adrian Chapman, and Sisi Zhou. This work has been supported by the IBM-UChicago Quantum Collaboration, under Agreement No. MAS000364, with access to the fleet of IBM Quantum computers. 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), NSF (OMA-2137642, OSI-2326767, CCF-2312755, OSI-2426975), and the Packard Foundation (2020-71479). L. E. F. acknowledges funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant Agreement No. 955479 (MOQS—Molecular Quantum Simulations).

Co-leading authors listed in alphabetical order.

DATA AVAILABILITY

The data and code supporting the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: PROOF FOR SELF-CONSISTENT ERROR MITIGATION

In this section, we give a rigorous proof for Theorem 1. For this purpose, we will first review the standard PEC procedure, prove its correctness, and then generalize to self-consistent PEC.

Let us first specify the model assumptions. For an n -qubit system, we consider the following set of operations and their noisy implementation.

1. Initialization: $|0\rangle\langle 0| \mapsto \tilde{\rho}_0 = \Lambda_S(|0\rangle\langle 0|)$.
2. Computational-basis measurement: $\{|b\rangle\langle b|\}_{b \in \{0,1\}^n} \mapsto \{\tilde{E}_b = \Lambda_M(|b\rangle\langle b|)\}_{b \in \{0,1\}^n}$.
3. Layer of single-qubit unitary: $\mathcal{U} = \bigotimes_{k=1}^n \mathcal{U}_k$, implemented without noise.
4. Layer of multi-qubit Clifford: $\mathcal{G} \mapsto \tilde{\mathcal{G}} = \mathcal{G} \circ \Lambda_G$, for all $\mathcal{G}$ from a finite set $\mathfrak{G}$.

Here, we further assume Λ_G are $\mathcal{G}$ -dependent Pauli channels, and Λ_S, Λ_M are generalized depolarizing channels (i.e., λ_a only depends on $\text{pt}(a)$). We use Λ to denote the collection of all noise channels. Furthermore, we assume all the Pauli eigenvalues are strictly positive. All these assumptions are experimentally justified via randomized compiling [8]. We also allow these Pauli channels to come from any quasi-local ansatzes, as introduced in the main text.

Though we only define the noise channel associated with the computational-basis measurement, since we assume single-qubit gates to be noiseless and Λ_M to be invariant under single-qubit rotation, we can effectively estimate any observable O up to the Λ_M , i.e., $\tilde{O} = \Lambda^M(O)$.

a. Standard PEC. Suppose we want to estimate the expectation value of an observable O at the output state of a quantum circuit. Denote the ideal value by

$$o = \langle\langle O | \mathcal{G}_T \mathcal{U}_T \cdots \mathcal{G}_1 \mathcal{U}_1 | \rho_0 \rangle\rangle. \quad (\text{A1})$$

Here, $\mathcal{U}_j$'s are layers of (possibly non-Clifford) single-qubit gates, and $\mathcal{G}_j$'s are layers of multi-qubit Clifford gates from $\mathfrak{G}$. Because of noise, a direct execution of the above gate sequence will instead give

$$\begin{aligned} o^{(\text{noisy})} &= \langle\langle \tilde{O} | \tilde{\mathcal{G}}_T \mathcal{U}_T \cdots \tilde{\mathcal{G}}_1 \mathcal{U}_1 | \tilde{\rho}_0 \rangle\rangle \\ &= \langle\langle O | \Lambda_M \mathcal{G}_T \Lambda_{\mathcal{G}_T} \mathcal{U}_T \cdots \mathcal{G}_1 \Lambda_{\mathcal{G}_1} \mathcal{U}_1 \Lambda_S | \rho_0 \rangle\rangle. \end{aligned} \quad (\text{A2})$$

To retrieve the ideal value, a naive idea is to cancel out all noise channels Λ by implementing Λ^{-1} . For any Pauli channel $\Lambda = \sum_b \lambda_b |P_b\rangle\langle P_b|$, its inverse is $\Lambda^{-1} = \sum_b \lambda_b^{-1} |P_b\rangle\langle P_b|$, which can be expressed as

$$\Lambda^{-1}(\rho) = \sum_{a \in \mathcal{P}^n} p_a^* P_a \rho P_a, \quad \text{where } p_a^* = \frac{1}{4^n} \sum_{b \in \mathcal{P}^n} (-1)^{\langle a, b \rangle} \lambda_b^{-1}. \quad (\text{A3})$$

This is a Pauli diagonal map that is not necessarily completely positive (i.e., p_a^* can be negative). Consequently, it cannot be directly implemented as a quantum channel. Instead, one can rewrite it in the following form

$$\begin{aligned} \Lambda^{-1}(\rho) &= \sum_a \left(\sum_b |p_b^*| \right) \frac{\text{sgn}_a |p_a^*|}{\sum_b |p_b^*|} P_a \rho P_a \\ &= \sum_a \gamma \text{sgn}_a q_a P_a \rho P_a, \end{aligned} \quad (\text{A4})$$

where sgn_a is the sign of p_a^* , $\gamma = \sum_b |p_b^*|$, and $q_a = |p_a^*|/\gamma$. Note that $\{q_a\}$ forms a probability distribution over $\mathcal{P}^n$. Thus, by sampling Pauli operator P_a according to $\{q_a\}$ and multiplying γsgn_a in classical post-processing, one can implement Λ^{-1} in expectation. This is the core idea of PEC.

Concretely, consider the following steps of *standard PEC* (which has assumed all noise channels are known *a priori*):

1. Randomly sample $a_0 \sim \{q_{a_0}^S\}$, $a_j \sim \{q_{a_j}^{\mathcal{G}_j}\}_{j=1}^T$, $a_{T+1} \sim \{q_{a_{T+1}}^M\}$.
2. Implement and measure the following expectation value

$$E_a = \langle\langle \tilde{O} | \mathcal{P}_{a_{T+1}} \tilde{\mathcal{G}}_T \mathcal{P}_{a_T} \mathcal{U}_T \cdots \tilde{\mathcal{G}}_1 \mathcal{P}_{a_1} \mathcal{U}_1 \mathcal{P}_{a_0} | \tilde{\rho}_0 \rangle\rangle \quad (\text{A5})$$

where $\mathcal{P}_a(\rho) = P_a \rho P_a$.

3. Define the PEC estimator as

$$\hat{o}^{(\text{PEC})} = E_a \cdot \prod_{j=0}^{T+1} \gamma_j \text{sgn}_{a_j}. \quad (\text{A6})$$

where γ_j and sgn_{a_j} are with respect to the j th Pauli noise channel.

The following proposition shows the correctness of PEC.
Proposition 2. Given that one knows Λ exactly, the standard PEC estimator $\hat{o}^{(\text{PEC})}$ is an unbiased estimator for o .
Proof.

$$\begin{aligned} \mathbb{E} \hat{o}^{(\text{PEC})} &= \sum_a E_a \cdot \prod_{j=0}^{T+1} \gamma_j \text{sgn}_{a_j} q_{a_j} = \sum_a E_a \cdot \prod_{j=0}^{T+1} p_{a_j}^* \\ &= \sum_a \langle\langle \tilde{O} | p_{a_{T+1}}^{*,M} \mathcal{P}_{a_{T+1}} \tilde{\mathcal{G}}_T p_{a_T}^{*,\mathcal{G}_T} \mathcal{P}_{a_T} \mathcal{U}_T \cdots \tilde{\mathcal{G}}_1 p_{a_1}^{*,\mathcal{G}_1} \mathcal{P}_{a_1} \mathcal{U}_1 p_{a_0}^{*,S} \mathcal{P}_{a_0} | \tilde{\rho}_0 \rangle\rangle = \langle\langle \tilde{O} | \Lambda_M^{-1} \tilde{\mathcal{G}}_T \Lambda_{\mathcal{G}_T}^{-1} \mathcal{U}_T \cdots \tilde{\mathcal{G}}_1 \Lambda_{\mathcal{G}_1}^{-1} \mathcal{U}_1 \Lambda_S^{-1} | \tilde{\rho}_0 \rangle\rangle \\ &= \langle\langle O | \mathcal{G}_T \mathcal{U}_T \cdots \mathcal{G}_1 \mathcal{U}_1 | \rho_0 \rangle\rangle = o. \end{aligned} \quad (\text{A7})$$

The second line is by the definition of q_a . ■

b. Self-consistent PEC. Formally, consider the following self-consistent PEC (SC-PEC) protocol. One first learns a set of noise parameters Λ_η that are gauge-equivalent to the true values Λ , meaning that the two noise models cannot be distinguished by any experiments constructed from the noisy gate set. Assuming the learning is exact for now. Use the superscript g to denote the learned noise channels. Construct our estimator using the following steps:

1. Randomly sample $a_0 \sim \{q_{a_0}^{\eta,S}\}$, $a_j \sim \{q_{a_j}^{\eta,\mathcal{G}_j}\}_{j=1}^T$, $a_{T+1} \sim \{q_{a_{T+1}}^{\eta,M}\}$.
2. Implement and measure the following expectation value

$$E_a = \langle\langle \tilde{O} | \mathcal{P}_{a_{T+1}} \tilde{\mathcal{G}}_T \mathcal{P}_{a_T} \mathcal{U}_T \cdots \tilde{\mathcal{G}}_1 \mathcal{P}_{a_1} \mathcal{U}_1 \mathcal{P}_{a_0} | \tilde{\rho}_0 \rangle\rangle, \quad (\text{A8})$$

which is formally the same as Eq. (A5).

3. Define the SC-PEC estimator as

$$\hat{o}^{(\text{SC-PEC})} = E_a \cdot \prod_{j=0}^{T+1} \gamma_j^\eta \text{sgn}_{a_j}^\eta. \quad (\text{A9})$$

where γ_j^η and $\text{sgn}_{a_j}^\eta$ are with respect to the j th learned Pauli noise channel (instead of the true noise channel).

The following proposition shows the correctness of SC-PEC.

Proposition 3. (Theorem 1 in main text). Given that one exactly knows a Λ_η that is gauge-equivalent to Λ , the SC-PEC estimator with respect to Λ_η is an unbiased estimator for o .

Proof. First note that E_a can be expanded as

$$E_a = \langle\langle O | \Lambda_M \mathcal{P}_{a_{T+1}} \mathcal{G}_T \Lambda_{\mathcal{G}_T} \mathcal{P}_{a_T} \mathcal{U}_T \cdots \mathcal{G}_1 \Lambda_{\mathcal{G}_1} \mathcal{P}_{a_1} \mathcal{U}_1 \mathcal{P}_{a_0} \Lambda_S | \rho_0 \rangle\rangle. \quad (\text{A10})$$

Since Λ and Λ_η are gauge-equivalent, replacing the former with the latter by definition does not change any expectation values from any experiments. We thus have,

$$E_a = \langle\langle O | \Lambda_{M,\eta} \mathcal{P}_{a_{T+1}} \mathcal{G}_T \Lambda_{\mathcal{G}_T,\eta} \mathcal{P}_{a_T} \mathcal{U}_T \cdots \mathcal{G}_1 \Lambda_{\mathcal{G}_1,\eta} \mathcal{P}_{a_1} \mathcal{U}_1 \mathcal{P}_{a_0} \Lambda_{S,\eta} | \rho_0 \rangle\rangle. \quad (\text{A11})$$

Then, following exactly the same argument as the proof of Proposition 2, one can obtain that

$$\mathbb{E} \hat{o}^{(\text{SC-PEC})} = \langle\langle O | \mathcal{G}_T \mathcal{U}_T \cdots \mathcal{G}_1 \mathcal{U}_1 | \rho_0 \rangle\rangle = o. \quad (\text{A12})$$

This completes the proof. ■

1. Generalizations to zero-noise extrapolation (ZNE)

The self-consistent formalism can be integrated with other quantum error mitigation methods besides PEC. Here, we discuss one example of self-consistent zero-noise extrapolation (ZNE). ZNE is a popular error mitigation method widely applied in experiments (e.g. [3]). At a high level, it works by intentionally amplifying all noise channels by a few different factors (i.e., $\Lambda \mapsto \Lambda^{1+\alpha}$), estimating the observable expectation values at different noise levels, and extrapolating to the zero-noise point using a certain fitting model. For the last step, one simple choice of models is single exponential decay, but it does not always yield provably unbiased estimators [3]. More complicated extrapolation models can be used at the price of potential numerical instability and difficulty in implementation.

The implementation of ZNE in Ref. [3] is based on probabilistic error amplification (PEA). That is, using random Pauli gates to implement a Pauli noise channel Λ^α [52], so as to realize arbitrary noise amplification. For this purpose, one needs to first characterize all the noise channels. We have seen that such characterization can only be done up to gauge degrees of freedom. Fortunately, once we learn a gauge-equivalent noise model Λ_η , PEC can be conducted self-consistently. To see this, consider again the following noisy expectation values

$$\begin{aligned} o^{(\text{noisy})} &= \langle\langle \tilde{O} | \tilde{\mathcal{G}}_T \mathcal{U}_T \cdots \tilde{\mathcal{G}}_1 \mathcal{U}_1 | \tilde{\rho}_0 \rangle\rangle \\ &= \langle\langle O | \Lambda_M \mathcal{G}_T \Lambda_{\mathcal{G}_T} \mathcal{U}_T \cdots \mathcal{G}_1 \Lambda_{\mathcal{G}_1} \mathcal{U}_1 \Lambda_S | \rho_0 \rangle\rangle. \end{aligned} \quad (\text{A13})$$

Suppose we exactly know a gauge-equivalent noise model Λ_η to the ground truth Λ . We can simply implement PEA with respect to the learned model to get

$$\begin{aligned} o^{(\text{PEA},\alpha)} &= \langle\langle O | (\Lambda_{M,\eta}^\alpha \Lambda_M) \mathcal{G}_T (\Lambda_{\mathcal{G}_T,\eta}^\alpha) \mathcal{U}_T \cdots \mathcal{G}_1 (\Lambda_{\mathcal{G}_1,\eta}^\alpha) \mathcal{U}_1 (\Lambda_{S,\eta}^\alpha \Lambda_S) | \rho_0 \rangle\rangle \\ &= \langle\langle O | \Lambda_{M,\eta}^{1+\alpha} \mathcal{G}_T \Lambda_{\mathcal{G}_T,\eta}^{1+\alpha} \mathcal{U}_T \cdots \mathcal{G}_1 \Lambda_{\mathcal{G}_1,\eta}^{1+\alpha} \mathcal{U}_1 \Lambda_{S,\eta}^{1+\alpha} | \rho_0 \rangle\rangle \\ &= \langle\langle O | \Lambda_M^{1+\alpha} \mathcal{G}_T \Lambda_{\mathcal{G}_T}^{1+\alpha} \mathcal{U}_T \cdots \mathcal{G}_1 \Lambda_{\mathcal{G}_1}^{1+\alpha} \mathcal{U}_1 \Lambda_S^{1+\alpha} | \rho_0 \rangle\rangle. \end{aligned} \quad (\text{A14})$$

In the first line, we implement Λ_η^α using random Pauli gates. The second line uses the definition of gauge-equivalent noise model to replace all Λ with Λ_η without changing the expectation values. The last line says this value will be gauge-independent. To see this, recall from the main text that that all gauges in a Pauli noise model can be parameterized as [21]:

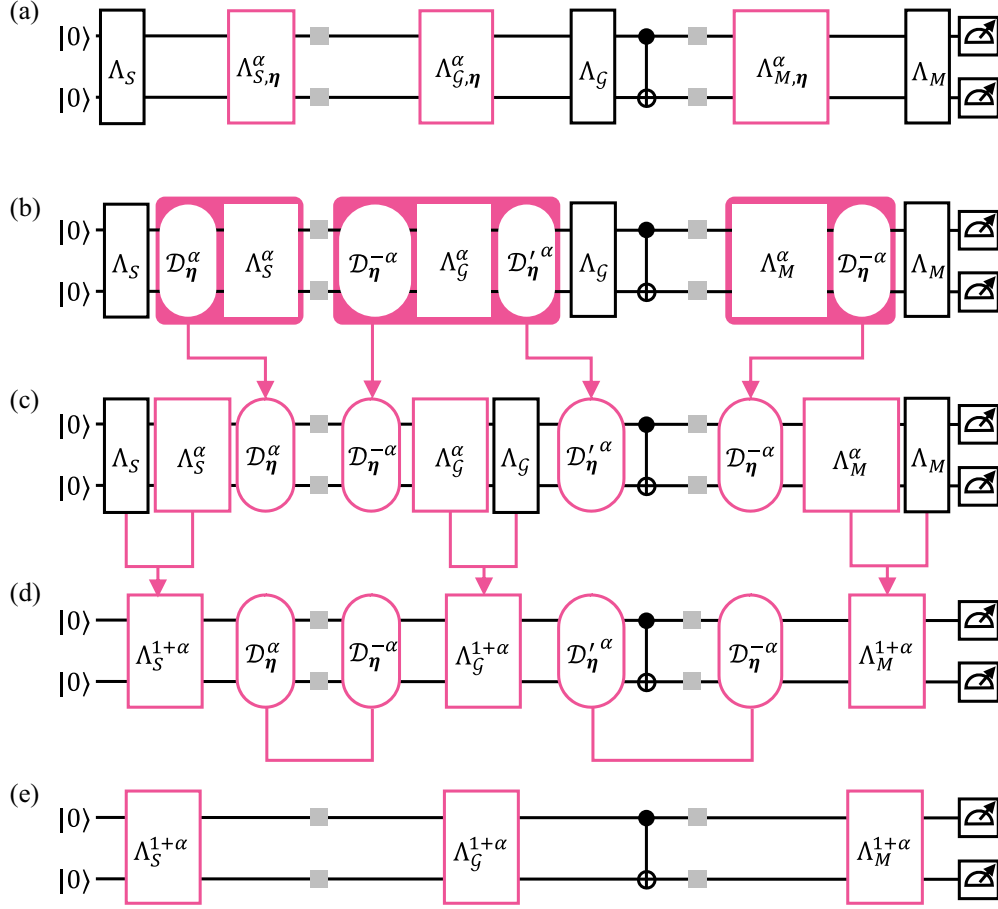


FIG. 9. Graphical proof of self-consistent ZNE. The only difference from Fig. 4 is that we implement Λ^α for general α instead of Λ_η^{-1} as in PEC. (a) Implementation of Λ_η^α via random Pauli gates. (b) Rewrite the noise channels using the ground truth and the gauges. (c) Commute the noise channels through. Note that both the noises and the gauges are Pauli diagonal maps and thus commute with one another. (d) Merge the noise channels. (e) Cancel the gauge maps. Recall that $\mathcal{D}_\eta'$ is defined according to $\mathcal{G} \circ \mathcal{D}_\eta' = \mathcal{D}_\eta \circ \mathcal{G}$.

$$\Lambda^S \mapsto \Lambda_\eta^S = \mathcal{D}_\eta \circ \Lambda^S, \quad (\text{A15})$$

$$\Lambda^M \mapsto \Lambda_\eta^M = \Lambda^M \circ \mathcal{D}_\eta^{-1}, \quad (\text{A16})$$

$$\Lambda^G \mapsto \Lambda_\eta^G = \mathcal{D}_\eta' \circ \Lambda^G \circ \mathcal{D}_\eta^{-1}. \quad (\text{A17})$$

Further note that all the above are Pauli diagonal maps that commute with each other. We thus have, for example, $(\mathcal{D}_\eta' \circ \Lambda^G \circ \mathcal{D}_\eta^{-1})^{1+\alpha} = \mathcal{D}_\eta'^{1+\alpha} \circ \Lambda^{G1+\alpha} \circ \mathcal{D}_\eta^{-(1+\alpha)}$. Putting these back into Eq. (A14) and canceling the gauge maps yield the last line. This can also be seen via a graphical proof as in Fig. 9, similar to our SC-PEC proof in the main text. In conclusion, we show that PEA (and thus ZNE) can be conducted self-consistently within our framework. In particular, the noise-amplified observable expectation values $\phi^{(\text{PEA}, \alpha)}$ are gauge-independent. Still, different choice of gauge leads to different PEA implementations, which may in turn lead to different variance and mitigation cost. Conducting gauge optimization for ZNE is an interesting open problem for future exploration.

APPENDIX B: HOW TO LEARN SELF-CONSISTENTLY

1. Details of the quasi-local model

In this section, we provide additional details about the quasi-local Pauli noise model. Let us first introduce the notion of factor sets. Let Ω be a subset of $2^{[n]}$, i.e., the power set of $[n] = \{1, 2, \dots, n\}$. We call Ω a factor set if for every $\kappa \in \Omega$, every subset of κ also belongs to Ω . An exemplary factor set on $n = 3$ qubits is given by $\Omega = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}\}$. For any nontrivial $P_a \in \mathcal{P}^n$, we say $a \sim \Omega$ if the Pauli support of a belongs to Ω . The set of all nontrivial Pauli operators given by Ω is denoted by

$$\mathcal{K} = \{a \sim \Omega : \forall a \in \mathcal{P}^n, a \neq 0\}. \quad (\text{B1})$$

In the above example, $XYI \in \mathcal{K}$ while $ZIX \notin \mathcal{K}$.

Recall that a Pauli channel is Ω -local if it can be expressed as

$$\Lambda(\cdot) = \bigcirc_{a \in \mathcal{K}} (\omega_a P_a(\cdot) P_a + (1 - \omega_a)(\cdot)), \quad (\text{B2})$$

with $\omega_a < 1/2$ and we define $\tau_a = -\log(1 - 2\omega_a)$. For any $\mathcal{K}$ defined via Eq. (B1) with a valid factor set Ω , the following relations are known [Eq. (12) in the main text],

$$x_a = \sum_{b \in \mathcal{K}} \langle a, b \rangle \tau_b. \quad (\text{B3})$$

The inverse transformation will become clear in a moment.

For the convenience of discussion, let us introduce another equivalent parameterization of Ω -local Pauli channels. For any two Pauli $a, b \in \mathcal{P}^n$, we write $a \triangleleft b$ if $\text{supp}(a) \subseteq \text{supp}(b)$ and that a, b commutes at every qubit. For example, $XIYI \triangleleft XZYZI$, while $XXZI \not\triangleleft YYYZ$.

Define $\mathbf{r} = \{r_a\}_{a \in \mathcal{K}}$ according to

$$\begin{cases} x_a = \sum_{b \in \mathcal{K}: b \triangleleft a} r_b, & \forall a \in \mathcal{P}^n, a \neq 0, \\ r_b = \sum_{a \in \mathcal{K}: a \triangleleft b} (-1)^{|b|-|a|} x_a, & \forall b \in \mathcal{K}. \end{cases} \quad (\text{B4})$$

This is known as the Möbius transform [36]. We again note that $\mathcal{K}$ must be defined via a valid factor set for the above to hold. $\mathbf{r}$ is referred to as the reduced parameter in Ref. [21]. The main advantages of using $\mathbf{r}$ are that every log eigenvalues $x_a = -\log \lambda_a$ can be very intuitively expressed in terms of $\mathbf{r}$ using the above equations. Thus, we will use $\mathbf{r}$ when discussing experimental design.

Finally, it can be verified that the following transformation from $\mathbf{r}$ to $\boldsymbol{\tau}$ holds. Note again that $\mathcal{K}$ needs to be defined via a valid factor set:

$$\tau_b = \sum_{c \in \mathcal{K}: \text{supp}(b) \subseteq \text{supp}(c)} \frac{-2}{4^{|c|}} (-1)^{\langle b, c \rangle} r_c, \quad (\text{B5})$$

where $\text{supp}(b)$ denote the Pauli support of b , i.e., the set of indexes i where $b_i \neq I$. Combined with Eq. (B4), one can obtain the inverse transformation from $\mathbf{x}$ to $\boldsymbol{\tau}$. All of these relations can be verified by direct substitution. They can also be found in [Ref. [21], Appendix E].

2. Constructing the design matrix

Recall from Sec. II A that the design matrix F encodes all experimental measurements. The matrix represents a linear map between the vector of observables $\mathbf{b}$ and $\mathbf{x}$, a vector of all (log) fidelities which varies in size depending on the gate set, the number of qubits, and the underlying locality of the noise. For convenience, we reproduce the key expression here:

$$\mathbf{b} = F\mathbf{x}, \quad (\text{B6})$$

where $b_j = -\log \langle \tilde{o}_j \rangle$ is the (log) expectation value of the j -th experiment.

For the 2Q experiments shown in Fig. 5 of the main text, it was not necessary to learn *all* noise parameters if the observable being mitigated is restricted to a certain type—in that case the Z -only observables. This *restricted* set of noise parameters was sufficient and complete as seen in Fig. 10(b), where Z -only observables on $n = 2$ qubits for depth-0, depth-1, and depth-2 (or more depth-even experiments) can saturate all learnable degrees of freedom subject to the remaining $2^n - 1$ gauge degrees of freedom.

We use this opportunity to describe the same analysis in a more practical noise parameterization written as $\mathbf{r}$, where the transformation $\mathbf{x} = M\mathbf{r}$ was defined earlier as the Möbius transformation seen in Eq. (B4). We show how the design matrix in the $\mathbf{r}$ basis, F' , is only slightly modified [Fig. 10(d)] without any change, in this special case, to the number of parameters in the noise model ($|\mathbf{x}| = |\mathbf{r}|$). In this more convenient basis, the design matrix F' can be seen to be complete as long as the matrix rank of F' is equal to $|\mathbf{r}| - (2^n - 1)$ for a general noise model, or $|\mathbf{r}| - n$ for two-local noise model that only admits single-qubit gauge transformation (e.g. Fig. 3) [21].

For larger system sizes n even with a *restricted* noise model, the number of learning experiments not only depends on the Hamming weight of the target observable, but also grows rapidly with the system size itself. In the case of the largest GHZ state we prepared on $n = 21$ qubits, the number of gauge parameters is, in theory, as large as $2^{21} - 1$. However, because we took advantage of the fact that the target circuit is a Clifford circuit, whose final observable could be classically back-propagated, we focused our experiments exclusively on learning those noise fidelities $\mathbf{x}$ which contributed to corrupting the target observable $\langle X^n \rangle$ (See Fig. 11 for an illustration of this procedure on one of the two template layers). To be exact, the $n = 21$ GHZ state required: a single, depth-0 experiment for SPAM, 7 depth-1 experiments, 7 depth-even experiments for each depths-even circuits of 2, 4, and 8. In total, 29 learning experiments informed the 56 observables needed to unambiguously infer the 46 fidelity terms in $\mathbf{x}$ (See Table I). The inferred noise model was used to predict $\langle X^n \rangle$ and compared against the experimentally measured value for the target circuit (Fig. 6). Although we performed this analysis in the $\mathbf{x}$ basis (as opposed to the $\mathbf{r}$ basis), we verified that the design matrix F was complete by observing that $\text{rank}(F)$, 34, and the number of SPAM bases, 12, add up to the total number of unknown fidelities $|\mathbf{x}| = 46$. This same procedure was used, with overlapping experiments where possible, for all the GHZ system sizes from $n = 3$ to $n = 21$.

To move beyond *restricted* noise models, we needed to impose locality in the noise so that the number of noise parameters did not continue to grow exponentially in system size. For this task, we utilized the design principle

(a) $\mathbf{b} = F\mathbf{x}$, e.g.

(b)

$$\begin{aligned}
 & -\log \begin{pmatrix} \langle ZI \rangle \\ \langle IZ \rangle \\ \langle ZZ \rangle \end{pmatrix}_{\text{depth-0}} \\
 & -\log \begin{pmatrix} \langle ZI \rangle \\ \langle IZ \rangle \\ \langle ZZ \rangle \end{pmatrix}_{\text{depth-1}} \\
 & -\log \begin{pmatrix} \langle ZI \rangle \\ \langle IZ \rangle \\ \langle ZZ \rangle \end{pmatrix}_{\text{depth-2}}
 \end{aligned}
 =
 \begin{pmatrix}
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
 1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1
 \end{pmatrix}_{9 \times 9}
 \begin{pmatrix}
 x_{10}^S \\
 x_{01}^S \\
 x_{11}^S \\
 x_{10}^M \\
 x_{01}^M \\
 x_{11}^M \\
 x_{ZI} \\
 x_{IZ} \\
 x_{ZZ}
 \end{pmatrix}_9$$

(c) $\mathbf{b} = F'\mathbf{r}$, where $F' = FM$, e.g.

(d)

$$\begin{aligned}
 & -\log \begin{pmatrix} \langle ZI \rangle \\ \langle IZ \rangle \\ \langle ZZ \rangle \end{pmatrix}_{\text{depth-0}} \\
 & -\log \begin{pmatrix} \langle ZI \rangle \\ \langle IZ \rangle \\ \langle ZZ \rangle \end{pmatrix}_{\text{depth-1}} \\
 & -\log \begin{pmatrix} \langle ZI \rangle \\ \langle IZ \rangle \\ \langle ZZ \rangle \end{pmatrix}_{\text{depth-2}}
 \end{aligned}
 =
 \begin{pmatrix}
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\
 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 & 1 \\
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1
 \end{pmatrix}_{9 \times 9}
 \begin{pmatrix}
 r_0^S \\
 r_1^S \\
 r_{01}^S \\
 r_0^M \\
 r_1^M \\
 r_{01}^M \\
 r_{Z0} \\
 r_{Z1} \\
 r_{Z0Z1}
 \end{pmatrix}_9$$

FIG. 10. (a) The general expression used to infer the (log) fidelity parameters of the noise model, $\mathbf{x}$, based on experimental observables $\mathbf{b}$. The experiments are linear combinations of state preparation, gates, and measurements fidelities, and are encoded into the design matrix F . (b) For example, we show the design of 2Q experiments limited to observables containing only Pauli Z terms. In this restricted noise model, there are only 9 noise terms ($|\mathbf{x}| = 9$), which can be inferred from depth-0, depth-1, and depth-2 (or more generally, a series of depth-even) experiments. The matrix rank of F in this case is 6, which is $|\mathbf{x}| = 9 - \dim(\text{gauge}) = 9 - 3$, as expected. (c) Although we present the design matrix F in the $\mathbf{x}$ fidelity basis within the theory section of the main text and also use it for the 2Q and GHZ experiments, we also draw attention to a definition in the $\mathbf{r}$ or noise parameter basis yielding F' , where $\mathbf{x} = M\mathbf{r}$, and M is the Möbius transformation defined in Eq. (B4). Defining F' in the $\mathbf{r}$ basis allows us to keep the number of noise parameters polynomial, as opposed to exponential, in number when we impose a quasi-local model for larger system sizes. (d) For example, we also present the F' in the $\mathbf{r}$ basis for the 2Q experiments here.

outlined in Ref. [21] and also briefly discussed throughout the sections above. Unlike the previous two examples, knowledge of the target observable was not used to inform the creation of the design matrix F' (in this case $\mathbf{r}$ basis)—instead, we conduct a *complete* learning of the quasi-local noise model. To measure the 9108 rows of observables for estimating all 2576 noise parameters in $\mathbf{r}$, we needed a total of 1 circuit for SPAM, 17 circuits for each template layer at depth-1, and 9 circuits for each template layer for multiple depth-even values (e.g. 4, 12, and 24). The explicit input and output bases can be found in Table II, and the additional details in Table III.

To characterize all the learnable parameters to additive precision it suffices to only perform a single, depth-0 SPAM experiment *and* additional depth-1 experiments for each layer. The depth-1 experiments involve the preparation of a Pauli eigenstate, a single application of the layer, terminated by a measurement in a Pauli basis that can be different than the initial one.

However, in the low-error regime, it is desirable to learn the parameters with multiplicative precision, which means the estimates can be improved with repeated applications of the gates. Therefore, we augment these experiments with additional even-depth experiments involving the

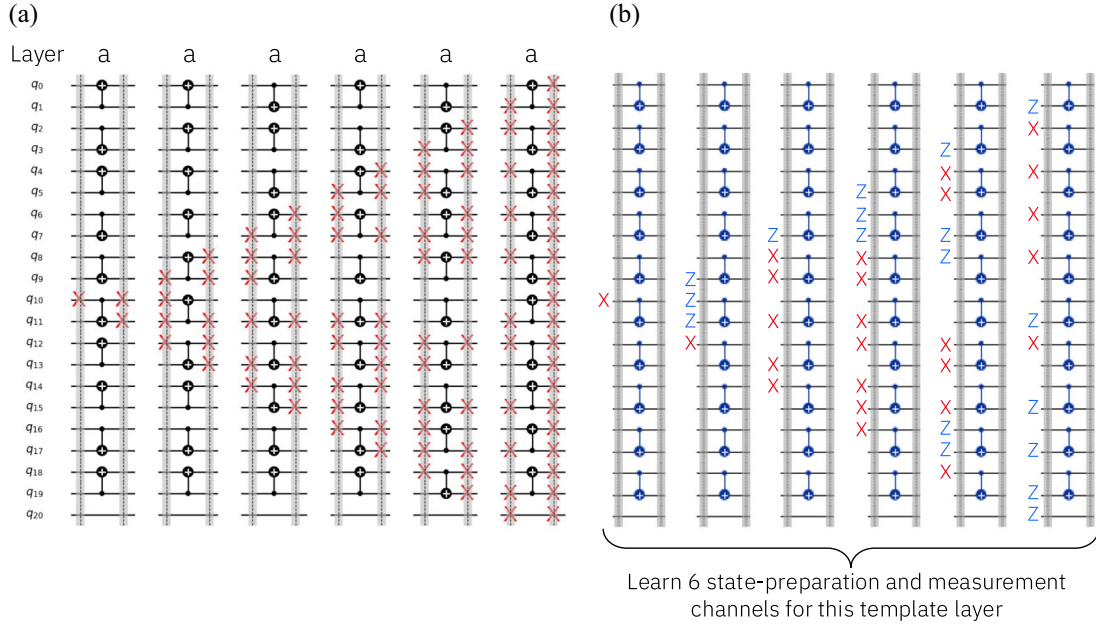


FIG. 11. For one of the two template layers a and b used for preparing the $n = 21$ GHZ state, we looked at those Pauli eigenvalues in (a) which contribute to the $\langle X^n \rangle$ observable, relying on the fact that this is a Clifford circuit and thus back-propagation was possible. (b) Since the template layer a may not have the CNOT in the same directions as the seen in the target circuit, the input and output bases must be properly accounted for after pre- and post-pending single-qubit Hadamard gates around the gate, thus the conversion from X -only input to some Z -type Paulis. For this template, six state preparation and measurement channels were needed. We performed the same task for template layer b , and grouped these experiments, commuting-wise, after appending the preparation and measurement bases into a single string of length $2n$. This allowed us to reduce the number of preparation and learning experiments down to 7, depth-1, and 7 depth-even experiments to invert the design matrix F of dimensions (56×46) .

preparation of a Pauli eigenstate, an even number of applications of the layer, and measurements in the *same* Pauli basis for a local two-qubit basis (9 experiments per depth per layer). Additionally, when the input and output Paulis commute qubit-wise, the corresponding experiments can be combined. With these considerations, we reduced the total circuit count for both layers from 54 to 34 for the depth-1 experiments, and from 38 to 18 for the depth-even experiments (See Table II for the exact input and output bases).

In our learning experiments, we measured depths of 4, 12, and 24 for the depth-even learning experiments. We emphasize that the number of experiments we have designed does not depend on the number of qubits, or the size of the qubit ring as long it is a multiple of four. Finally, to ensure numerical stability in inferring the model parameters from the *logarithm* of the measured expectation values $\langle O \rangle$, the largest depth of these learning experiments need to be smaller than the inverse of the typical gate error rate, e.g. for gate errors of $\approx 1\%$, circuit depths should be less than ≈ 100 ; also, learning circuits sampled with enough repetitions such that statistical fluctuations $\sigma_{\langle O \rangle}$ of the measured observables are much smaller than the measured outcomes $\langle O \rangle$.

APPENDIX C: EXPERIMENTAL DETAILS

In this section we give further details on the experiments presented in the main text. We refer to these as the two-qubit experiment from Sec. IV A, the GHZ preparation experiment in Sec. IV B, and the ring circuit experiment in Sec. IV C. We summarize the main differences between these experiments in Table III.

For all executed circuits, we employed uniform Pauli twirling of the respective two-qubit gate layers to suppress coherent errors and justify the assumption of a Pauli noise channel. That is, we ran several instances of circuits, known as “twirls,” that implement the same global unitary but differ in their single-qubit gate layers. Despite this randomized circuit compilation overhead, we maintained kHz sampling rates by making use of a recently introduced parametric circuit compilation and parameter binding pipeline facilitated by the `Sampler` primitive within the IBM Qiskit runtime service [69], and the by the `samplomatic` package [70] for the PEC experiments. Moreover, we symmetrized the noise channel of the readout by also twirling measurements through random insertion of Pauli $\hat{X}$ or $\hat{I}$ gates (sampled uniformly) prior to the readout [19]. Nonetheless, the overhead of running different twirling and measurement configurations remains non-negligible, which is why we collected multiple measurements (“shots”) for each twirled circuit (See Table III).

TABLE I. To mitigate the $n = 21$ GHZ experiment, we required a total of 56 learning experiments (rows of the design matrix) containing a total of 46 unique fidelity parameters (columns of the design matrix). The fidelity parameters λ with superscripts “s” denote state-preparation, “m” for measurement, “1” for layer-1, and “2” for layer-2. The Pauli eigenvalues are the subscripts ordered from qubits 1 to 21, from left to right.

Count	Depth/Type	Experiment
1	Depth-even	$2 * x^1_{IIIIIIIXXIXIXIXIXI} + 2 * x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
2	Depth-1	$x^1_{IXXIXIXIXIXIXIXIXI} + x^m_{IXXIXIXIXIXIXIXIXI} + x^s_{IXXIXIXIXIXIXIXIXI}$
3	Depth-even	$x^1_{IIIIIIIXXIXIXIXIXI} + x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
4	Depth-1	$x^1_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
5	Depth-even	$2 * x^2_{ZZZIZIZIXIXIXIXIXI} + 2 * x^2_{ZZZZZZXIXIXIXIXIXI} + x^m_{ZZZIZIZIXIXIXIXIXI} + x^s_{ZZZIZIZIXIXIXIXIXI}$
6	Depth-even	$2 * x^2_{IIIIIIIXXIXIXIXIXI} + 2 * x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
7	Depth-even	$2 * x^2_{IIIXXIXIXIXIXIXIXI} + 2 * x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
8	Depth-even	$2 * x^2_{IIIIIIIXXIXIXIXIXI} + 2 * x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
9	Depth-even	$2 * x^1_{IIIXXIXIXIXIXIXIXI} + 2 * x^1_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
10	Depth-even	$2 * x^1_{IIIIIIIXXIXIXIXIXI} + 2 * x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
11	Depth-even	$4 * x^1_{IIIXXIXIXIXIXIXIXI} + 4 * x^1_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
12	Depth-even	$2 * x^1_{IIIXXIXIXIXIXIXIXI} + 2 * x^1_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
13	Depth-even	$4 * x^2_{IIIIIIIXXIXIXIXIXI} + 4 * x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
14	Depth-even	$4 * x^2_{IIIXXIXIXIXIXIXIXI} + 4 * x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
15	Depth-even	$x^2_{IIIIIIIXXIXIXIXIXI} + x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
16	Depth-even	$x^2_{IIIXXIXIXIXIXIXIXI} + x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
17	Depth-even	$x^2_{IIIIIIIXXIXIXIXIXI} + x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
18	Depth-even	$4 * x^1_{IIIIIIIXXIXIXIXIXI} + 4 * x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
19	Depth-1	$x^2_{ZZZIZIZIXIXIXIXIXI} + x^m_{ZZZIZIZIXIXIXIXIXI} + x^s_{ZZZIZIZIXIXIXIXIXI}$
20	Depth-1	$x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
21	Depth-1	$x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
22	Depth-1	$x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
23	Depth-1	$x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
24	Depth-1	$x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
25	Depth-even	$x^2_{ZZZIZIZIXIXIXIXIXI} + x^2_{ZZZZZZXIXIXIXIXIXI} + x^m_{ZZZIZIZIXIXIXIXIXI} + x^s_{ZZZIZIZIXIXIXIXIXI}$
26	Depth-even	$4 * x^2_{ZZZIZIZIXIXIXIXIXI} + 4 * x^2_{ZZZZZZXIXIXIXIXIXI} + x^m_{ZZZIZIZIXIXIXIXIXI} + x^s_{ZZZIZIZIXIXIXIXIXI}$
27	Depth-even	$x^2_{IIIIIIIXXIXIXIXIXI} + x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
28	Depth-even	$x^2_{IIIXXIXIXIXIXIXIXI} + x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
29	Depth-even	$4 * x^2_{IIIIIIIXXIXIXIXIXI} + 4 * x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
30	Depth-even	$4 * x^2_{IIIXXIXIXIXIXIXIXI} + 4 * x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
31	Depth-even	$4 * x^2_{IIIIIIIXXIXIXIXIXI} + 4 * x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
32	Depth-even	$2 * x^2_{IIIIIIIXXIXIXIXIXI} + 2 * x^2_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
33	Depth-even	$2 * x^2_{IIIXXIXIXIXIXIXIXI} + 2 * x^2_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
34	Depth-even	$x^1_{IIIXXIXIXIXIXIXIXI} + x^1_{IIIXXIXIXIXIXIXIXI} + x^m_{IIIXXIXIXIXIXIXIXI} + x^s_{IIIXXIXIXIXIXIXIXI}$
35	Depth-1	$x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
36	Depth-1	$x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
37	Depth-even	$x^1_{IXXIXIXIXIXIXIXIXI} + x^1_{IXXIXIXIXIXIXIXIXI} + x^m_{IXXIXIXIXIXIXIXIXI} + x^s_{IXXIXIXIXIXIXIXIXI}$
38	Depth-even	$2 * x^1_{IXXIXIXIXIXIXIXIXI} + 2 * x^1_{IXXIXIXIXIXIXIXIXI} + x^m_{IXXIXIXIXIXIXIXIXI} + x^s_{IXXIXIXIXIXIXIXIXI}$
39	Depth-even	$4 * x^1_{IIIIIIIXXIXIXIXIXI} + 4 * x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
40	Depth-even	$4 * x^1_{IIIIIIIXXIXIXIXIXI} + 4 * x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
41	Depth-even	$4 * x^1_{IXXIXIXIXIXIXIXIXI} + 4 * x^1_{IXXIXIXIXIXIXIXIXI} + x^m_{IXXIXIXIXIXIXIXIXI} + x^s_{IXXIXIXIXIXIXIXIXI}$
42	Depth-even	$x^1_{IIIIIIIXXIXIXIXIXI} + x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
43	Depth-even	$x^1_{IIIIIIIXXIXIXIXIXI} + x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
44	Depth-1	$x^1_{IIIIIIIXXIXIXIXIXI} + x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
45	SPAM	$x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
46	SPAM	$x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
47	SPAM	$x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
48	SPAM	$x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$
49	SPAM	$x^m_{IIIIIIIXXIXIXIXIXI} + x^s_{IIIIIIIXXIXIXIXIXI}$

(Table continued)

TABLE I. (*Continued*)

Count	Depth/Type	Experiment
50	SPAM	$x^m_{ZZZZZZZZZZZZZZZZZZZZ} + x^s_{ZZZZZZZZZZZZZZZZZZZZ}$
51	SPAM	$x^m_{ZZIZIZIZIZIZIZIZIZIZI} + x^s_{ZZIZIZIZIZIZIZIZIZIZI}$
52	SPAM	$x^m_{IIIIIIIZZIZIZIIIIII} + x^s_{IIIIIIIZZIZIZIIIIII}$
53	SPAM	$x^m_{IIIZZIZZZIZIZIZIIII} + x^s_{IIIZZIZZZIZIZIZIIII}$
54	SPAM	$x^m_{IIZZZIZZZIZIZIZIZII} + x^s_{IIZZZIZZZIZIZIZIZII}$
55	SPAM	$x^m_{IIIIIIIIIZZIZIIIIIIII} + x^s_{IIIIIIIIIZZIZIIIIIIII}$
56	SPAM	$x^m_{IIIIIZZIZIZIIIZZIZIIII} + x^s_{IIIIIZZIZIZIIIZZIZIIII}$

TABLE II. This table shows the input- and output-bases for the learning circuits needed for SPAM (1), for depth-1 observables (17 per layer) and depth-even observables (9 per layer per depth) necessary for the design matrix F . Those output bases in gray can be measured simultaneously using the output bases at the top of each cell. This assumes 12 qubits on a closed ring. For more qubits, which also must be a multiple of four and on a closed ring, the bases are repeated with a period of four. For other qubit topologies, such as lines or lattices, new designs must be undertaken.

Count	Layer 0		Layer a		Layer b	
	Input	Output	Input	Output	Input	Output
SPAM	<i>ZZZZZZZZZZZZ</i>	<i>ZZZZZZZZZZZZ</i>				
Depth-1						
1			YZYZYZYZYZYZ	XYXYXYXYXYXY XYXYXYXYXYXY	ZYZYZYZYZYZYZ	YXYXYXYXYXYXY YXYXYXYXYXYXY
2		YYYYYYYYYYYY	XZXZXZXZXZXZ	XZXZXZXZXZXZ	ZXZXZXZXZXZX	YYYYYYYYYYYY YYYYYYYYYYYY
3		XZXZXZXZXZXZ	YYYYYYYYYYYY	YYYYYYYYYYYY	YYYYYYYYYYYY	ZXZXZXZXZXZX ZXZXZXZXZXZX
4		XYXYXYXYXYXY	YZYZYZYZYZYZ	YZYZYZYZYZYZ	YXYXYXYXYXYXY	ZYZYZYZYZYZYZ ZYZYZYZYZYZYZ
5		ZYXXZYXXZYXX	ZYXXZYXXZYXX	IYXIIYXIIYXI ZIIIXZIIIXZII	XZYXXZYXXZYX	XZYXXZYXXZYX IIYXIIYXIIYX XZIIIXZIIIXZII
6		XXZYXXZYXXZY	XXZYXXZYXXZY	XIIYXIIYXIIY IXZIIIXZIIIXZI	YXXZYXXZYXXZ	YXXZYXXZYXXZ YXIIYXIIYXII IIXZIIIXZIIIXZ
7		ZYYXZYXXZYXX	IYYIIYYIIYYI	IYYIIYYIIYYI	XZYXXZYXXZYX	IIYYIIYYIIYY IIYYIIYYIIYY
8		YXZYXXZYXXZY	YIIYYIIYYIIY	YIIYYIIYYIIY	YXXZYXXZYXXZ	YIIYYIIYYIIY YIIYYIIYYIIY
9		ZZXXZZXXZZXX	IZXIIZXIIZXI	IZXIIZXIIZXI	XZZXXZZXXZZX	IIZXIIZXIIZX IIZXIIZXIIZX
10		XXZZXXZZXXZZ	XIIZXIIZXIIZ	XIIZXIIZXIIZ	ZXXZZXXZZXXZ	ZXIIZXIIZXII ZXIIZXIIZXII
11		ZZYXZZYXZZYX	IZYIIZYIIZYI	IZYIIZYIIZYI	XZZYXZZYXZZY	IIZYIIZYIIZY IIZYIIZYIIZY
12		YXZZYXZZYXZZ	YIIZYIIZYIIZ	YIIZYIIZYIIZ	ZYXZZYXZZYXZ	ZYIIZYIIZYII ZYIIZYIIZYII
13		XXXXXXXXXXXX	XXXXXXXXXXXX XXXXXXXXXXXX IXXIIXXIIXXI XIIIXIIXXIIIX	XXXXXXXXXXXX XXXXXXXXXXXX	XXXXXXXXXXXX	XXXXXXXXXXXX XXXXXXXXXXXX IIXXIIXXIIIX XXIIIXIIXXII
14		YXYXYXYXYXYX	YXYXYXYXYXYX YXYXYXYXYXYX IXYIIXYIIXYI YIIXYIIXYIIX	YXYXYXYXYXYX YXYXYXYXYXYX	YXYXYXYXYXYX	YXYXYXYXYXYX YXYXYXYXYXYX IIXYIIXYIIXY XYIIXYIIXYII

(Table continued)

TABLE II. (Continued)

Count	Layer 0		Layer a		Layer b	
	Input	Output	Input	Output	Input	Output
SPAM	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ				
Depth-1						
15			ZXZXZXZXZXZX	ZXZXZXZXZXZX	XZXZXZXZXZXZ	XZXZXZXZXZXZ
16			ZYZYZYZYZYZY	ZYZYZYZYZYZY	YZZYZYZYZYZY	YZZYZYZYZYZY
			IYZIYZIYZI	IYZIYZIYZI	IYZIYZIYZI	IYZIYZIYZI
			ZIYZIYZIYZ	ZIYZIYZIYZ	ZIYZIYZIYZ	ZIYZIYZIYZ
17			ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
			ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ
			IZZIIZZIIZZI	IZZIIZZIIZZI	IZZIIZZIIZZI	IZZIIZZIIZZI
			ZIIZZIIZZIIZ	ZIIZZIIZZIIZ	ZIIZZIIZZIIZ	ZIIZZIIZZIIZ
Depth-even						
1			XXXXXXXXXXXXX	XXXXXXXXXXXXX	XXXXXXXXXXXXX	XXXXXXXXXXXXX
2			XYXYXYXYXYXY	XYXYXYXYXYXY	XYXYXYXYXYXY	XYXYXYXYXYXY
3			XZXZXZXZXZXZ	XZXZXZXZXZXZ	XZXZXZXZXZXZ	XZXZXZXZXZXZ
4			YXYXYXYXYXYX	YXYXYXYXYXYX	YXYXYXYXYXYX	YXYXYXYXYXYX
5			YYYYYYYYYYYYY	YYYYYYYYYYYYY	YYYYYYYYYYYYY	YYYYYYYYYYYYY
6			YZZYZYZYZYZY	YZZYZYZYZYZY	YZZYZYZYZYZY	YZZYZYZYZYZY
7			ZXZXZXZXZXZX	ZXZXZXZXZXZX	ZXZXZXZXZXZX	ZXZXZXZXZXZX
8			ZYZYZYZYZYZY	ZYZYZYZYZYZY	ZYZYZYZYZYZY	ZYZYZYZYZYZY
9			ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ	ZZZZZZZZZZZZ

For each experiment, we learned both a “inconsistent” noise model and a self-consistent noise model. The inconsistent models derive from the learning theory originally established in Ref. [9]: For each noisy layer, we implemented a given number of even-depth learning circuits for a basis of Paulis as specified in Table III. In this context, SPAM errors were dealt with independently from gate noise following the technique from Ref. [19] known also as *twirled readout error extinction* (TRES). That is, the noisy expectation value of an observable O was divided by an estimate of $\langle 0|O|0 \rangle$ in a prepare- $|0\rangle$ circuit (under measurement twirling). The noisy estimate of $\langle 0|O|0 \rangle$ was performed with the same number of twirls and shots per twirl as stated in Table III. Finally, the set of learning circuits for the self-consistent noise models comprises the same even-depth learning circuits used for the inconsistent model as well as additional depth-one learning circuits for the respective Pauli basis of the model.

1. Details on two-qubit experiments

Whereas we discussed the details of the learning circuits earlier in Appendix B 2, here we will focus on the two-qubit target circuit we examined. For this *restricted* model experiment limited to Z-only observables, we prepared all learning circuits in the $|00\rangle$ state, but used that learned noise model to mitigate the outcomes of a circuit prepared

in the $|11\rangle$ state. In this manner, by comparing the experimentally measured outcomes for the observables $\langle IZ \rangle$, $\langle ZI \rangle$, and $\langle ZZ \rangle$, against those predicted by the learned noise model, we were able to identify an improved bias, of up to 4% for a single pair of qubits up to depth-32, and also an improved bias of 0.97% across six qubits on *ibm_auckland*, a 27-qubit device (See Fig. 12). This comprised of a total of 144 learning experiments, and 144 mitigation experiments taken over the course of a day.

2. Details on GHZ-preparation experiments

In Sec. IV B of the main text and above, we only focused our discussion on $n = 5$ or $n = 21$ GHZ states. However, the full data set involved all odd-sized GHZ states between $n = 3$ and $n = 21$, inclusive. We used 21 physical qubits on a line: 12, 17, 30, 31, 32, 36, 51, 50, 49, 55, 68, 69, 70, 74, 89, 88, 87, 93, 106, 105, 104. We repeated the experiment a total of 7 times, interleaving the learning experiments and the target experiments over the course of a day (18 hours) immediately after the system *ibm_strasbourg* was calibrated. The standard deviations shown in Fig. 6 were taken over the 7 experimental runs.

3. Details on 92-qubit ring experiments

Here we detail the circuit and observables for the experiments presented in Sec. IV C of the main text.

TABLE III. Details for the four experimental results, with increasing complexity, presented in the same order as the main text. To transform the noise of both the learning circuits and the target circuits into Pauli noise, a number of logically equivalent circuits, referred to as “twirls” was implemented. For each twirl, the circuit was sampled with a given number of shots, where each shot lasted for approximately 1 millisecond. Depending on the target circuit, the number of even-layers used to infer the symmetric noise model ranged in depths from 8 to 40. Finally, the complexity of the noise model increased with the size of circuits, where for the two-qubit experiment the model Pauli basis was restricted to those that impacted the mitigated observables whereas the 20-qubit and 92-qubit experiment involved all one- and two-local Paulis. This also meant the design matrix for the self-consistent noise model grew not only in the number of rows (corresponding to the number of learned observables), but also the number of columns (corresponding to the noise parameters) which grow from 9 to 2576. We emphasize that the final two columns represent the *scalable* approach, where the total number of learning circuits remains fixed no matter the size of the system; as discussed in the main text, this is because the noise is assumed to be two-local and thus long-range noise terms are not being considered. For examples of design matrices, see Fig. 10 for the 2Q column, and Table I for the GHZ column. The larger design matrices are not shown, but can be reconstructed using code provided in Ref. [21].

Experiment	Single CNOT (see Sec. IV A)	GHZ preparation (see Sec. IV B)	Ring circuit (see Sec. IV C)	Non-Clifford circuit (see Sec. IV D)
Number of qubits	2	21	92	20
Observables	ZZ, ZI	$X^{\otimes n}$	$Z_i, i \in \{0, \dots, 91\}$	$Z_i, i \in \{0, \dots, 19\}$
Number of twirls	250	100	100	128 (learning) 100.000 (PEC)
Shots per twirl	200	256	150	64 (learning) 1 (PEC)
Even-depth learning layers for symmetric model	$d = \{2, 4, 6, \dots, 32\}$	$d = \{2, 4, 8\}$	$d = \{4, 12, 24\}$	$d = \{4, 10, 20, 40\}$
Model Pauli basis	Restricted to Paulis relevant for observable	Restricted to Paulis relevant for observable	All one- and two-local Paulis	All one- and two-local Paulis
Design matrix dimensions of self-consistent model	(9, 9)	(56, 46)	(9108, 2576)	(2349, 521)
Model locality assumption	None	None	Two-local	Two-local

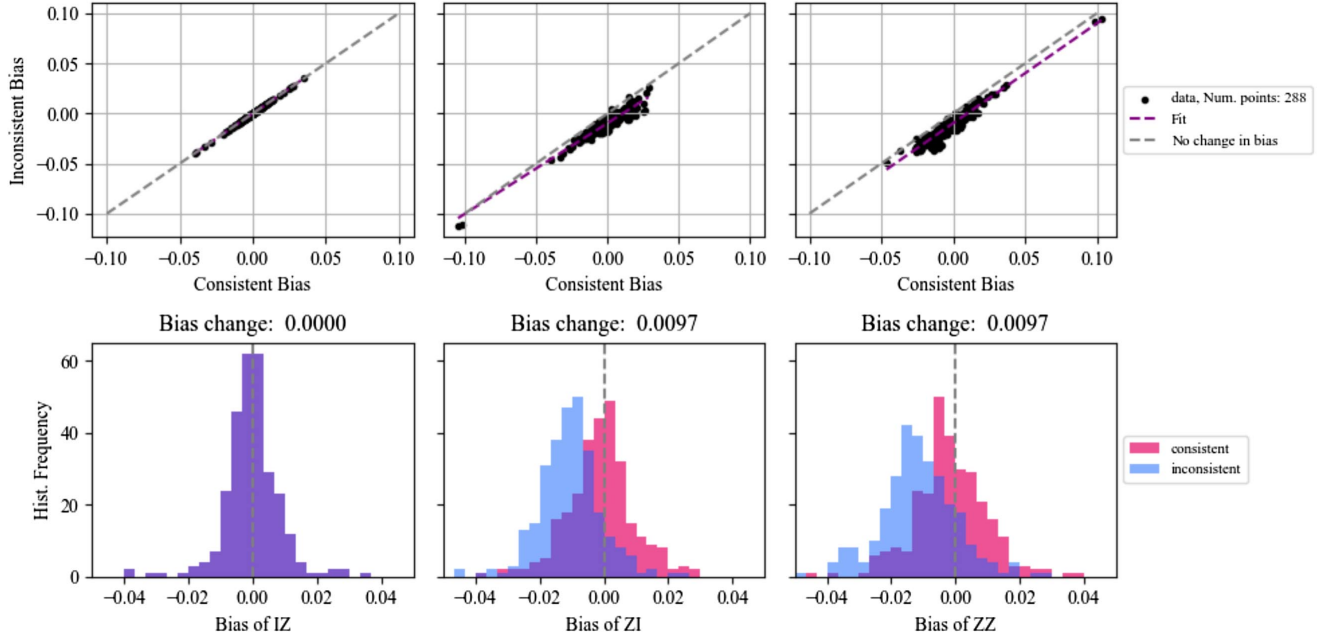


FIG. 12. Two-qubit experiments across 27-qubit device, *ibm_auckland*. (**Top**) Comparison of bias in mitigated values of the consistent versus the inconsistent noise models for the $\langle IZ \rangle$, $\langle ZI \rangle$, and $\langle ZZ \rangle$ observables. (**Bottom**) Histogram distributions of the biases for the consistent (red) and the inconsistent (blue) noise models for the $\langle IZ \rangle$, $\langle ZI \rangle$, and $\langle ZZ \rangle$ observables.

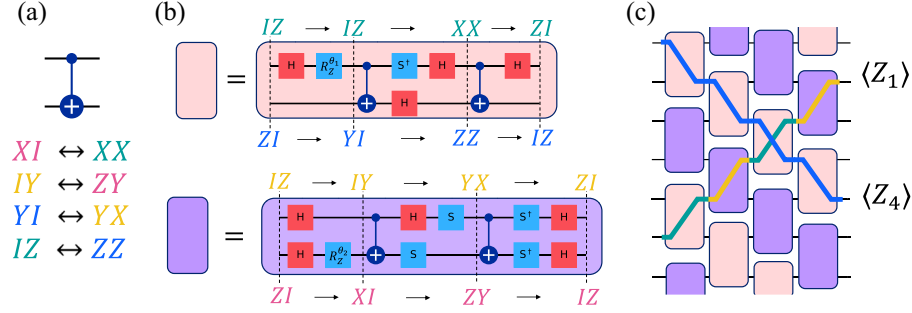


FIG. 13. Circuit for ring experiments with weight-1 observables. (a) A single CNOT gate has four conjugate Pauli fidelity pairs that change pattern under conjugation. (b) Two qubit circuit blocks that transform $ZI \leftrightarrow IZ$, such that the noisy Pauli fidelities that affect each observable (before the unitary part of the gate) originate from different conjugate pairs, as indicated by color. The indicated Pauli paths arise for the R_z rotation angles $\theta_1 = \pi/2$ and $\theta_2 = -\pi/2$, which we use for the Clifford circuit experiments from Sec. IV C. (c) Arranging the two-qubit blocks from (b) in an alternating pattern, each single-qubit Z observable propagates in a staircase shape, such that every degenerate fidelity from each CNOT gate shown in (a) is probed by one observable.

This experiment was designed with the aim to probe all degenerate fidelity pairs of the CNOT gates of a one-dimensional, closed loop of qubits. For a single CNOT gate, there are four degenerate cycles of conjugate Pauli pairs that change their pattern under conjugation with the CNOT gate. These are $IZ \leftrightarrow ZZ$, $XI \leftrightarrow XX$, $ZY \leftrightarrow IY$, and $YX \leftrightarrow YI$ [20] [See Fig. 13(a)]. We start by designing two different two-qubit blocks with two noisy CNOT gates each, such that all single-qubit Z observables are sensitive to two Pauli fidelities that originate from different degenerate cycles. These two-qubit blocks are shown in Fig. 13(a). Note that they shift the support of the IZ observable to the other qubit. Hence, when arranging the blocks in the pattern shown in Fig. 13(c), each Z-observable propagates in a “staircase”-like trajectory. After four layers of the alternating pattern (eight total layers of CNOTs), every fidelity split of each participating CNOT connection is probed by one observable. There are only two unique layers of CNOT gates, as the two-qubit blocks only differ in their single-qubit gate structure.

With this construction, we ensure that the Z observables in this experiment are maximally sensitive to asymmetries in the gate noise fidelities. Moreover, every observable is affected by the state preparation noise from one qubit and the measurement noise of a different qubit. However, traditional approaches of readout error mitigation (used for the “symmetric model” throughout this work) are sensitive to the state preparation error of the final, measured qubit [19]. Hence, these observables expose flaws in the symmetric model when state preparation errors are not uniform across the ring of qubits.

The measured observables generally showed better agreement with the self-consistent noise model than with the symmetric noise model [See Fig. 7(e)]. However, for most observables, a residual bias remained also under the self-consistent noise model. This is an indication that there are error sources present in the device which even the consistent model does not accurately account for.

Candidates for such errors could be leakage out of the qubit subspace, temporal drifts of the noise model between the learning circuits and the target circuits, remaining coherent errors, or non-nearest-neighbor correlated noise sources. Finally, few individual observables in Fig. 7(e) (e.g. qubit index 57) show a larger bias under the self-consistent model than the symmetric model. We note that this occurred predominantly when there were severe outliers in the individual Pauli or SPAM fidelities affecting the respective observables. This could be caused, e.g., by the presence of two-level systems (TLS). These lead to strong fluctuations in the noise parameters on short time scales, to which the self-consistent learning protocol is particularly vulnerable due its dependence on depth-one circuits. We thus expect that our learning protocol will further benefit from recent techniques to stabilize the noise [48].

4. Details on PEC experiment

We built on the benchmark circuits introduced in Appendix C 3 for our proof-of-principle demonstration of gauge-optimized PEC presented in Sec. IV D. Specifically, we used versions of the circuits presented in Fig. 13 obtained by choosing $\theta_1 = -\pi/3$ and $\theta_2 = 2\pi/3$. In this setting, the Pauli paths of the measured single-qubit Z observables become more complicated than in the Clifford case; they generally split into two branches with every two-qubit block, but still probe the degenerate fidelity pairs of each CNOT and are sensitive to differences in SPAM errors. Instead of running the circuits on a closed loop of qubits with periodic boundary conditions as in Sec. IV C, for this experiment we chose to run the experiments on an open-ended 1d string of qubits to further optimize the error rates of the chosen qubit layout. In this setting, the two-qubit blocks that are only shown halfway in Fig. 13(c) were removed from the circuit. Hence, at a circuit size of 20 qubits, the two different layers of parallel CNOT gates have 10 and 9 gates,

respectively. At a depth of 10 layers of alternating blocks, this led to 190 CNOT gates in total as shown in Fig. 8(b).

APPENDIX D: GAUGE OPTIMIZATION

1. Difference in γ between two strategies

Our initial naïve attempt at optimizing the gauge involved using the Moore-Penrose pseudo-inverse to recover $\hat{\mathbf{r}}_0 = \mathbf{F}'^+ \mathbf{b}$, which is effectively a least-squares minimization problem of the form $\epsilon = \|\mathbf{F}' \mathbf{r}_0 - \mathbf{b}\|_2$ [71]. This is followed by a second optimization step:

$$\min_{\boldsymbol{\eta} \in \mathbb{R}^n} \left\{ \sum_{i=1}^{24n} \max([A_r^T(\mathbf{r}_0 + \mathbf{S}^\dagger \boldsymbol{\eta})]^{(i)}, 0) \right\} \quad (\text{D1})$$

where the k th column of the matrix $\mathbf{S}^\dagger$ is one of n vectors $\mathbf{y}_k$ in the nullspace of the design matrix. In other words, $\mathbf{S}^\dagger$ converts each gauge parameter $\boldsymbol{\eta}_k$ into a vector in the noise parameter space $\mathcal{X}$ such that the residual errors ϵ between the model and measured outcomes remains unchanged: $\mathbf{F}' \hat{\mathbf{r}}_0 - \mathbf{b} = \mathbf{F}'(\hat{\mathbf{r}}_0 + \mathbf{S}^\dagger \boldsymbol{\eta}) - \mathbf{b}$. The summand in the optimization problem limits all $24n$ elements of the $\boldsymbol{\tau}$ vector to be positive definite before being summed, element-wise, together. The optimized $\hat{\mathbf{r}}_*^{\text{two-step}}$ is then an offset of $\hat{\mathbf{r}}_0$: $\hat{\mathbf{r}}_0 + \mathbf{S}^\dagger \boldsymbol{\eta}^*$, where finding $\boldsymbol{\eta}^*$ involves solving the convex optimization problem in Eq. (D1).

However, benchmarking on the noise model data from the 92-qubit experiment from Sec. IV C, we found that this

two-step approach, which started with a $\gamma \approx 264$ without any optimization step (no steps beyond calculating $\hat{\mathbf{r}}_0 = \mathbf{F}'^+ \mathbf{b}$), yielded much higher overheads than even the γ inferred from a non-negative least-squares fit based on previous approaches, where $\gamma \approx 20.97$ [9]. Thus, the optimization procedure we outlined in Sec. III B yielded a slightly higher residual error ($\epsilon = 329.39$) but at a significantly lower $\gamma = 16.33$ (See Fig. 14).

2. Gauge-optimized model for PEC

The above analysis illustrates that there is a bias-variance tradeoff when choosing the gauge such that the PEC overhead γ is minimized. As shown in Fig. 14, a small concession in the fitting error ϵ (which may introduce a bias due to a worse fit of the noise model to the learning data) can significantly reduce the variance of a PEC estimator. For our demonstration of gauge-optimized PEC in Sec. IV D, we navigated this tradeoff by exploring a region of possible ϵ that are within the shot noise fluctuations of the learning data, as detailed in the following.

First, we note that the vector $\hat{\mathbf{b}}$ of measured log-fidelities in Eq. (9) is subject to shot noise on the individual expectation values $\tilde{d}_j$. Let us denote the statistical errors on these expectation values as $\delta \tilde{d}_j$. We obtained the “default” gauge $\hat{\mathbf{r}}^0$ of the learned noise model by solving Eq. (16) directly, where the sum over layers was taken such that each mitigated layer enters the expression as many times as it occurs in the circuit, i.e., once for the state preparation layer and ten times for each of the two CNOT layers. This way, the cost function directly represents the γ -value of the full circuit, for which we obtain $\gamma^0 = 48.7$. Note that we mitigate readout errors in post-processing by dividing the PEC estimators through the readout fidelities

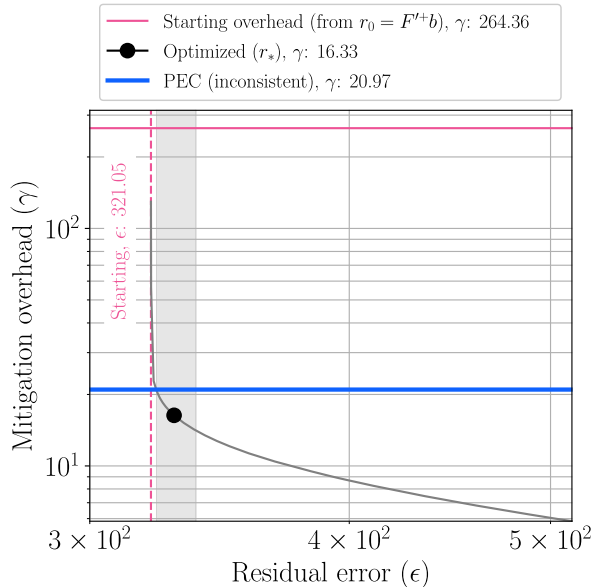


FIG. 14. Gauge optimization for the noise model of the 92-qubit experiments (Sec. IV C), where the target residual error $\epsilon = \|\mathbf{F}' \mathbf{r} - \mathbf{b}\|_2$ can be varied. The black point was chosen as the optimal trade-off between residual error and overhead minimization, with the shaded gray region indicating the range of $\mathbf{r}$ values we chose from.

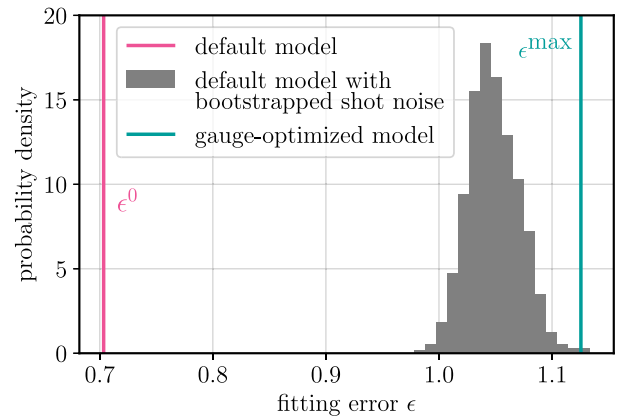


FIG. 15. Illustration of the choice of the residual error ϵ_{max} for the gauge-optimized PEC experiment presented in Sec. IV D, where ϵ_{max} was chosen such that it lies just within the observed fitting errors of the original model when accounting for shot noise in the learning data through bootstrapping.

corresponding to the measured observables, in the spirit of the TREX method.

The resulting fitting error ϵ^0 establishes a baseline shown as the leftmost vertical line in Fig. 15. Next, we computed bootstrapped vectors $\hat{\mathbf{b}}_{\text{bs}}$ by sampling new sets of values for $\tilde{o}_j$ drawing from appropriate Bernoulli statistics (since all observables are Paulis with possible measurement outcomes of ± 1) with mean $\tilde{o}_j$ and standard deviation $\delta\tilde{o}_j$. We repeated this process 1000 times to obtain fittings errors $\|F'\hat{\mathbf{r}}^0 - \hat{\mathbf{b}}_{\text{bs}}\|$. The distribution of these errors is shown as a histogram in Fig. 15. For the gauge-optimized model, we chose the maximum allowed fitting error as $\epsilon^{\text{max}} = 1.6\epsilon^0$, which lies within the tail of the distribution of the bootstrapped values, see Fig. 15. This reduced the γ value to 3.7, which corresponds to an improvement in sampling overhead from 2372 for the default gauge to only 13.7, an improvement by a factor of 173.

3. Computational efficiency of gauge optimization

We now discuss the computational efficiency of the gauge optimization problem. For convenience, we restate the gauge optimization problem:

$$\begin{aligned} \min_{\hat{\mathbf{r}}} & \left\{ \sum_{a \in \mathcal{K}, \text{layer}} \max(\hat{\tau}_a^{\text{layer}}, 0) \right\}, \\ \text{s.t. } & \|F'\hat{\mathbf{r}} - \hat{\mathbf{b}}\|_2 \leq \epsilon. \end{aligned} \quad (\text{D2})$$

Here, the summation is over all layers (including both SPAM and gates) of noise and all Pauli operators consistent with our quasi-local noise ansatz (described by $\mathcal{K}$). For each layer, $\hat{\tau}$ can be obtained from $\hat{\mathbf{r}}$ via the linear transformation defined in Eq. (B5). We will assume there are polynomially many layers and $\mathcal{K}$ contains polynomially many Paulis (which is a reasonable requirement for a parameter-efficient noise ansatz, say nearest-neighbor two-local).

The above gauge optimization is a convex optimization problem. More specifically, it is a second-order cone programming (SOCP) problem. Indeed, it can be rewritten in the following equivalent form:

$$\begin{aligned} \min_{\hat{\mathbf{r}}, \mathbf{s}} & \sum_{i=1}^{|\boldsymbol{\tau}|} s_i, \\ \text{s.t. } & s_i \geq (L\hat{\mathbf{r}})_i, \\ & s_i \geq 0, \\ & \|F'\hat{\mathbf{r}} - \hat{\mathbf{b}}\|_2 \leq \epsilon. \end{aligned} \quad (\text{D3})$$

Here, we collect all τ_a^{layer} into one column vector $\boldsymbol{\tau}$ of length $|\boldsymbol{\tau}|$, which is completely determined by $\hat{\mathbf{r}}$ via a linear transformation we denote by $\boldsymbol{\tau} = L\hat{\mathbf{r}}$. The equivalence between the above two optimization problems is obvious,

and the latter is in a standard SOCP form (which consists of linear target functions and affine or 2-norm restrictions). Therefore, this problem is provably solvable to any given precision in polynomial time using the interior-point method [44].

In our experiments, we use CVXPY [43], a standard python-based convex optimization toolbox, to solve the gauge optimization problem. For our 92-qubit nearest-neighbor noise model experiments, the gauge optimization is solved within a few seconds on laptop, justifying its computational efficiency in practice.

-
- [1] S. P. Jordan, *Quantum algorithm zoo*, <https://quantumalgorithmzoo.org>.
 - [2] D. C. McKay, I. Hincks, E. J. Pritchett, M. Carroll, L. C. G. Govia, and S. T. Merkel, *Benchmarking quantum processor performance at scale*, [arXiv:2311.05933](https://arxiv.org/abs/2311.05933).
 - [3] 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 (London)* **618**, 500 (2023).
 - [4] 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).
 - [5] 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).
 - [6] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, *The future of quantum computing with superconducting qubits*, *J. Appl. Phys.* **132**, 160902 (2022).
 - [7] D. A. Lidar and T. A. Brun, *Introduction to decoherence and noise in open quantum systems*, in *Quantum Error Correction* (2013), p. 3.
 - [8] J. J. Wallman and J. Emerson, *Noise tailoring for scalable quantum computation via randomized compiling*, *Phys. Rev. A* **94**, 052325 (2016).
 - [9] 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).
 - [10] L. E. Fischer, M. Leahy, A. Eddins, N. Keenan, D. Ferracin, M. A. C. Rossi, Y. Kim, A. He, F. Pietracaprina, B. Sokolov, S. Dooley, Zán Zimborás, F. Tacchino, S. Maniscalco, J. Goold, G. García-Pérez, I. Tavernelli, A. Kandala, and S. N. Filippov, *Dynamical simulations of many-body quantum chaos on a quantum computer*, *Nat. Phys.* **22**, 302 (2026).
 - [11] K. Temme, S. Bravyi, and J. M. Gambetta, *Error mitigation for short-depth quantum circuits*, *Phys. Rev. Lett.* **119**, 180509 (2017).
 - [12] S. Endo, S. C. Benjamin, and Y. Li, *Practical quantum error mitigation for near-future applications*, *Phys. Rev. X* **8**, 031027 (2018).
 - [13] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, *Gate set tomography*, *Quantum* **5**, 557 (2021).

- [14] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, *Quantum computation with universal error mitigation on a superconducting quantum processor*, *Sci. Adv.* **5**, eaaw5686 (2019).
- [15] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, *Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system*, *Nat. Commun.* **11**, 587 (2020).
- [16] M. T. Madzik, S. Asaad, A. Youssry, B. Joecker, K. M. Rudinger, E. Nielsen, K. C. Young, T. J. Proctor, A. D. Baczewski, A. Laucht *et al.*, *Precision tomography of a three-qubit donor quantum processor in silicon*, *Nature (London)* **601**, 348 (2022).
- [17] 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).
- [18] A. Calzona, M. Papič, P. Figueroa-Romero, and A. Auer, *Multilayer cycle benchmarking for high-accuracy error characterization*, *Phys. Rev. Res.* **8**, 013111 (2026).
- [19] E. van den Berg, Z. K. Mineev, and K. Temme, *Model-free readout-error mitigation for quantum expectation values*, *Phys. Rev. A* **105**, 032620 (2022).
- [20] S. Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, and L. Jiang, *The learnability of Pauli noise*, *Nat. Commun.* **14**, 52 (2023).
- [21] S. Chen, Z. Zhang, L. Jiang, and S. T. Flammia, *Efficient self-consistent learning of gate set Pauli noise*, *PRX Quantum* **7**, 010305 (2026).
- [22] D. M. Greenberger, M. A. Horne, and A. Zeilinger, *Going beyond Bell's theorem*, in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos, Fundamental Theories of Physics Vol. 37 (Springer, Dordrecht, 1989), pp. 69–72, [10.1007/978-94-017-0849-4_10](https://doi.org/10.1007/978-94-017-0849-4_10).
- [23] 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, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, *Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor*, *Phys. Rev. X* **11**, 041039 (2021).
- [24] 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).
- [25] H.-Y. Huang, S. T. Flammia, and J. Preskill, *Foundations for learning from noisy quantum experiments*, [arXiv:2204.13691](https://arxiv.org/abs/2204.13691).
- [26] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, *A polynomial-time classical algorithm for noisy random circuit sampling*, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (2023), pp. 945–957.
- [27] 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](https://arxiv.org/abs/2303.17714).
- [28] S. T. Flammia and J. J. Wallman, *Efficient estimation of Pauli channels*, *ACM Trans. Quantum Comput.* **1**, 1 (2020).
- [29] E. van den Berg and P. Wocjan, *Techniques for learning sparse Pauli-Lindblad noise models*, *Quantum* **8**, 1556 (2024).
- [30] S. T. Flammia, *Averaged circuit eigenvalue sampling*, in *Proceedings of the 17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, *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, [10.4230/LIPICs.TQC.2022.4](https://doi.org/10.4230/LIPICs.TQC.2022.4).
- [31] 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).
- [32] E. T. Hockings, A. C. Doherty, and R. Harper, *Improving error suppression with noise-aware decoding*, [arXiv:2502.21044](https://arxiv.org/abs/2502.21044).
- [33] E. Pelaez, 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](https://arxiv.org/abs/2403.12857).
- [34] R. Harper, S. T. Flammia, and J. J. Wallman, *Efficient learning of quantum noise*, *Nat. Phys.* **16**, 1184 (2020).
- [35] R. Harper and S. T. Flammia, *Learning correlated noise in a 39-qubit quantum processor*, *PRX Quantum* **4**, 040311 (2023).
- [36] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, *Learning logical Pauli noise in quantum error correction*, *Phys. Rev. Lett.* **130**, 200601 (2023).
- [37] C. Rouzé and D. S. Franca, *Efficient learning of the structure and parameters of local Pauli noise channels*, [arXiv:2307.02959](https://arxiv.org/abs/2307.02959).
- [38] Y. Li and S. C. Benjamin, *Efficient variational quantum simulator incorporating active error minimization*, *Phys. Rev. X* **7**, 021050 (2017).
- [39] R. Haghshenas, E. Chertkov, M. Mills, W. Kadow, S.-H. Lin, Y.-H. Chen, C. Cade, I. Niesen, T. Begušić, M. S. Rudolph *et al.*, *Digital quantum magnetism on a trapped-ion quantum computer*, *Nature (London)* **653**, 56 (2026).
- [40] S. Filippov, M. Leahy, M. A. Rossi, and G. García-Pérez, *Scalable tensor-network error mitigation for near-term quantum computing*, [arXiv:2307.11740](https://arxiv.org/abs/2307.11740).
- [41] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, *Robust shadow estimation*, *PRX Quantum* **2**, 030348 (2021).
- [42] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, *Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit*, [arXiv:1310.4492](https://arxiv.org/abs/1310.4492).
- [43] S. Diamond and S. Boyd, *CVXPY: A Python-embedded modeling language for convex optimization*, *J. Mach. Learn. Res.* **17**, 1 (2016).
- [44] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, *Applications of second-order cone programming*, *Linear Algebra Appl.* **284**, 193 (1998).
- [45] N. Gavrielov, A. Ivrii, and S. Garion, *Linear circuit synthesis using weighted steiner trees*, *Quantum Inf. Comput.* **24**, 721 (2024).
- [46] N. Yoshioka, M. Amico, W. Kirby, P. Jurcevic, A. Dutt, B. Fuller, S. Garion, H. Haas, I. Hamamura, A. Ivrii *et al.*, *Krylov diagonalization of large many-body Hamiltonians on a quantum processor*, *Nat. Commun.* **16**, 5014 (2025).
- [47] L. C. G. Govia, S. Majumder, S. V. Barron, B. Mitchell, A. Seif, Y. Kim, C. J. Wood, E. J. Pritchett, S. T. Merkel, and D. C. McKay, *Bounding the systematic error in quantum*

- error mitigation due to model violation, *PRX Quantum* **6**, 010354 (2025).
- [48] Y. Kim, L. C. Govia, A. Dane, E. van den Berg, D. M. Zajac, B. Mitchell, Y. Liu, K. Balakrishnan, G. Keefe, A. Stabile *et al.*, *Error mitigation with stabilized noise in superconducting quantum processors*, *Nat. Commun.* **16**, 8439 (2025).
- [49] A. Seif, H. Liao, V. Tripathi, K. Krsulich, M. Malekakhlagh, M. Amico, P. Jurcevic, and A. Javadi-Abhari, *Suppressing correlated noise in quantum computers via context-aware compiling*, in *Proceedings of the 2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)*, Buenos Aires, Argentina (IEEE, 2024), pp. 310–324, [10.1109/ISCA59077.2024.00031](https://doi.org/10.1109/ISCA59077.2024.00031).
- [50] J. M. Gambetta, *Software for the era of quantum advantage*, in *IEEE Quantum Week Keynote*, Albuquerque, NM (2025).
- [51] S. V. Barron *et al.*, *Observable estimation in the absence of classical verification*, [arXiv:2607.25998](https://arxiv.org/abs/2607.25998).
- [52] A. Seif, M. Malekakhlagh, S. Majumder, and L. C. G. Govia, *Single snapshot non-Markovianity of Pauli channels*, [arXiv:2602.13145](https://arxiv.org/abs/2602.13145).
- [53] J. Kattemölle, B. Gulácsi, and G. Burkard, *Non-Markovianity induced by Pauli-twirling*, [arXiv:2602.08464](https://arxiv.org/abs/2602.08464).
- [54] D. Aharonov, O. Alberton, I. Arad, Y. Atia, E. Baïrey, Z. Brakerski, I. Cohen, O. Golan, I. Gurwich, O. Kenneth, E. Leviatan, N. H. Lindner, R. A. Melcer, A. Meyer, G. Schul, and M. Shutman, *On the importance of error mitigation for quantum computation*, [arXiv:2503.17243](https://arxiv.org/abs/2503.17243).
- [55] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill *et al.*, *Quantum advantage in learning from experiments*, *Science* **376**, 1182 (2022).
- [56] S. Chen, S. Zhou, A. Seif, and L. Jiang, *Quantum advantages for Pauli channel estimation*, *Phys. Rev. A* **105**, 032435 (2022).
- [57] A. Seif, S. Chen, S. Majumder, H. Liao, D. S. Wang, M. Malekakhlagh, A. Javadi-Abhari, L. Jiang, and Z. K. Mineev, *Entanglement-enhanced learning of quantum processes at scale*, [arXiv:2408.03376](https://arxiv.org/abs/2408.03376).
- [58] N. P. de Leon, K. M. Itoh, D. Kim, K. K. Mehta, T. E. Northup, H. Paik, B. S. Palmer, N. Samarth, S. Sangtawesin, and D. W. Steuerman, *Materials challenges and opportunities for quantum computing hardware*, *Science* **372**, eabb2823 (2021).
- [59] M. Malekakhlagh, A. Seif, D. Pizzuoli, L. C. G. Govia, and E. van den Berg, *Efficient Lindblad synthesis for noise model construction*, *npj Quantum Inf.* **11**, 191 (2025).
- [60] Z. Zhang, S. Chen, Y. Liu, and L. Jiang, *Generalized cycle benchmarking algorithm for characterizing midcircuit measurements*, *PRX Quantum* **6**, 010310 (2025).
- [61] J. Hines and T. Proctor, *Pauli noise learning for mid-circuit measurements*, *Phys. Rev. Lett.* **134**, 020602 (2025).
- [62] N. Tantivasadakarn, A. Vishwanath, and R. Verresen, *Hierarchy of topological order from finite-depth unitaries, measurement, and feedforward*, *PRX Quantum* **4**, 020339 (2023).
- [63] H. Buhrman, M. Folkertsma, B. Loff, and N. M. P. Neumann, *State preparation by shallow circuits using feed forward*, *Quantum* **8**, 1552 (2024).
- [64] R. S. Gupta, E. van den Berg, M. Takita, D. Ristè, K. Temme, and A. Kandala, *Probabilistic error cancellation for dynamic quantum circuits*, *Phys. Rev. A* **109**, 062617 (2024).
- [65] A. Deshpande, M. Hinsche, S. Najafi, K. Sharma, R. Sweke, and C. Zoufal, *Dynamic parameterized quantum circuits: Expressive and barren-plateau free*, [arXiv:2411.05760](https://arxiv.org/abs/2411.05760).
- [66] E. H. Chen, G.-Y. Zhu, R. Verresen, A. Seif, E. Bäumer, D. Layden, N. Tantivasadakarn, G. Zhu, S. Sheldon, A. Vishwanath *et al.*, *Nishimori transition across the error threshold for constant-depth quantum circuits*, *Nat. Phys.* **21**, 161 (2025).
- [67] E. Bäumer, V. Tripathi, D. S. Wang, P. Rall, E. H. Chen, S. Majumder, A. Seif, and Z. K. Mineev, *Efficient long-range entanglement using dynamic circuits*, *PRX Quantum* **5**, 030339 (2024).
- [68] J. Bausch, A. W. Senior, F. J. H. Heras, T. Edlich, A. Davies, M. Newman, C. Jones, K. Satzinger, M. Y. Niu, S. Blackwell, G. Holland, D. Kafri, J. Atalaya, C. Gidney, D. Hassabis, S. Boixo, H. Neven, and P. Kohli, *Learning high-accuracy error decoding for quantum processors*, *Nature (London)* **635**, 834 (2024).
- [69] 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](https://arxiv.org/abs/2405.08810).
- [70] I. Hincks, J. Skanes-Norman, S. Ferracin, and T. Armon, *Qiskit Samplomatic* (2025), <https://github.com/Qiskit/samplomatic>.
- [71] P. Virtanen *et al.* SciPy 1.0 Contributors, *SciPy 1.0: Fundamental algorithms for scientific computing in python*, *Nat. Methods* **17**, 261 (2020).