

Multilayer cycle benchmarking for high-accuracy error characterization

Alessio Calzona^{1,*}, Miha Papič^{1,2}, Pedro Figueroa-Romero¹ and Adrian Auer¹

¹*IQM Quantum Computers, Georg-Brauchle-Ring 23-25, Munich 80992, Germany*

²*Department of Physics and Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians-Universität München, Theresienstrasse 37, Munich 80333, Germany*



(Received 18 January 2025; accepted 4 December 2025; published 30 January 2026)

Accurate noise characterization is essential for reliable quantum computation. Effective Pauli noise models have emerged as powerful tools, offering a detailed description of the error processes with a manageable number of parameters, which guarantees the scalability of the characterization procedure. However, a fundamental limitation in the learnability of Pauli eigenvalues impedes a full high-accuracy characterization of both general and effective Pauli noise models, thereby restricting, e.g., the performance of noise-aware error mitigation techniques. We introduce multilayer cycle benchmarking (MLCB), an enhanced characterization protocol that improves the learnability associated with effective Pauli noise models by jointly analyzing multiple layers of Clifford gates. We show a simple experimental implementation and demonstrate that, in realistic scenarios, MLCB can reduce unlearnable noise degrees of freedom by up to 75%, improving the accuracy of sparse Pauli-Lindblad noise models and boosting the performance of error mitigation techniques like probabilistic error cancellation. Our results highlight MLCB as a scalable, practical tool for precise noise characterization and improved quantum computation.

DOI: [10.1103/4hbl-hp13](https://doi.org/10.1103/4hbl-hp13)

I. INTRODUCTION

A wide variety of characterization techniques, serving diverse purposes and spanning different ranges of applicability, have been developed to date. For instance, gate set tomography (GST) provides a comprehensive method to determine individual errors associated with executing sets of quantum operations [1], but its exponential scaling in sampling complexity limits its application to large systems [2]. On the other hand, holistic techniques like cross-entropy benchmarking [3,4] or scalable variants of randomized benchmarking [5–7] aim to benchmark large quantum processing unit (QPUs) by assessing their average quality.

Recently, considerable attention has been devoted to characterization schemes that lie between these extremes, offering greater detail than holistic techniques while still remaining scalable [8–12]. This trend is also driven by the need to implement noise-aware error mitigation techniques [8,13–15] and noise-aware decoding in the context of quantum error correction [16,17], both requiring detailed knowledge of the noise associated with the parallel execution of instructions on large QPUs. Achieving this without incurring the prohibitively high sampling complexity of GST typically requires implementing averaging techniques to recast the original noise into more manageable forms. Equally important is the development of

simpler, effective noise models that capture the most relevant features of the noise on the entire QPU while featuring a manageable number of parameters [8,11,18].

A prominent example of noise recasting techniques is Pauli twirling (PT) applied to Clifford gates, which twirls noise into Pauli noise [19–21]. The latter is a type of stochastic noise modeled by Pauli channels, where only errors associated with any Pauli operator occur with a given probability. Incidentally, this makes PT also a valuable error suppression tool, as it gets rid of a large portion of detrimental, coherent errors that get twirled into less detrimental, incoherent ones. This highlights the appeal and practical significance of characterizing Pauli noise [8,18]. Its relevance is further amplified by the possibility to develop and use effective and scalable versions of Pauli noise, where reasonable assumptions about noise locality can significantly reduce the number of parameters to be determined [8,9,18].

Despite these advantages, the characterization of gate-dependent Pauli noise remains fundamentally limited by Pauli learnability constraints, as demonstrated in Ref. [22] and confirmed experimentally [8,21,23]. As in GST [2,24], gauge freedom introduces ambiguities, making it impossible to determine certain parameters of Pauli channels independently from state preparation and measurement (SPAM) errors. Specifically, Ref. [22] shows that for a set of Clifford gates, the number of these unlearnable noise degrees of freedom (DOF) in Pauli channels scales exponentially with the number of qubits n . In this context, procedures based on cycle benchmarking (CB) [25,26] have emerged as convenient methods for accurately determining all learnable DOF of Pauli channels [22].

In this work, we specifically consider the characterization of scalable and effective Pauli noise models and present

*Contact author: alessio.calzona@meetiqm.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.

an enhanced version of CB called multilayer cycle benchmarking (MLCB), which significantly increases the number of learnable DOF by leveraging the constraints inherent in the effective noise model. Unlike standard CB implementations [8,18], MLCB simultaneously analyzes multiple sets of parallel Clifford gates, rather than characterizing each set in isolation. After providing a general framework for implementing MLCB, we focus on the characterization of the IQM Garnet™ [27] quantum processor, which features 20 superconducting qubits in a square topology. Our results show that MLCB can achieve a remarkable 75% reduction in the number of unlearnable DOF in practically relevant scenarios. Moreover, we experimentally highlight how MLCB-enabled learnable DOF differ significantly from estimates derived through conventional low-accuracy approaches. Additionally, we numerically show how this enhanced learnability offered by MLCB translates into more accurate characterization of effective noise models, focusing on the sparse Pauli-Lindblad (SPL) noise model introduced in Ref. [8] and widely used to inform noise-aware error mitigation strategies [8,14,15,23]. In this respect, we also investigate numerically how MLCB reduces residual biases in Probabilistic Error Cancellation (PEC), a prominent error mitigation technique that requires precise noise characterization to achieve optimal results [8,13,23,28].

The rest of the paper is organized as follows: Secs. II and III A provide a short overview of the established CB technique for Clifford gate layer characterization, discuss Pauli noise learnability constraints, and briefly review scalable Pauli noise models, with a focus on the SPL model. The MLCB framework is presented in Sec. IV, followed by its experimental testing on the IQM Garnet™ platform in Sec. V. Finally, Secs. VI and VII numerically illustrate how MLCB enhances characterization accuracy and improves the performance of noise-aware error mitigation techniques.

II. CHARACTERIZATION OF A CLIFFORD LAYER

A quantum circuit can be decomposed into a sequence of layers, where we define a layer as a set of instructions with nonoverlapping qubit supports executed in parallel. Characterizing the performance of entire layers, rather than just individual gates, is crucial for assessing errors in practical applications, especially due to effects like spatial crosstalk where operations on some qubits also affect their neighbors. This work focuses on characterizing the noise of a n -qubit Clifford layer C consisting of two-qubit entangling gates, whose execution typically represents the main source of errors. We model the noisy implementation of the layer C as a map $\mu(U_C)[\rho] := (\phi(U_C) \circ \tilde{\Lambda}_C)[\rho]$, where $\phi(U_C)[\rho] := U_C \rho U_C^\dagger$ is associated with the unitary operator $U_C \in \mathbb{U}(2^n)$ and $\tilde{\Lambda}_C$ is a completely positive and trace-preserving noisy map. The latter can be simplified into an effective Pauli channel

$$\Lambda_C[\rho] = \sum_{\alpha} p_{\alpha} P_{\alpha} \rho P_{\alpha}^{\dagger} \quad (P_{\alpha} \in \mathbb{P}_n = \{I, X, Y, Z\}^{\otimes n}), \quad (1)$$

by using PT, which requires sandwiching the layer of interest C with two layers of single-qubit Pauli gates [18–21]. As adjacent layers of single-qubit Pauli operators can then be

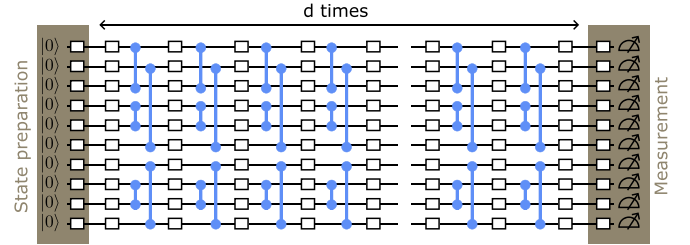


FIG. 1. Standard CB protocol: In the state preparation stage (left brown area), an eigenstate of a Pauli operator is prepared. The Clifford layer C subject to characterization (in this case, consisting of several parallel CZ gates shown in blue) is then implemented d times, alternating it with layers of single-qubit gates (white boxes) that are needed to implement RC. This can be interpreted as the implementation of d dressed Clifford layers. After a final layer of single-qubit gates that ensure that RC does not modify the unitary associated with the circuit, the measurement stage (right brown box), which also incorporates readout twirling, provides us with the expectation value of a Pauli operator (typically the same operator as considered in the state preparation stage).

compiled together [20], the Pauli map $\Lambda_C[\rho]$ can thus be understood as the noise affecting the dressed Clifford layer C , i.e., the layer C combined with a single layer of (random) Pauli gates. For more details about the twirling procedure, we refer the reader to Appendix A as well as to Refs. [18–20]. The task of characterizing the map $\Lambda_C[\rho]$ reduces to determining its Pauli eigenvalues f_{β} , associated with each Pauli operator P_{β} as $\Lambda_C[P_{\beta}] = f_{\beta} P_{\beta} \forall P_{\beta} \in \mathbb{P}_n$. Indeed, these Pauli eigenvalues, sometimes also referred to as Pauli fidelities [18,22,29], can be directly related to the error probabilities in Eq. (1) via the Walsh-Hadamard transform $p_{\alpha} = 2^{-n} \sum_{\beta} f_{\beta} (-1)^{\langle \alpha, \beta \rangle_{sp}}$, with $\langle \alpha, \beta \rangle_{sp}$ being the symplectic inner product of Pauli operators P_{α} and P_{β} , which is zero if they commute and one otherwise.

Products of Pauli eigenvalues can be efficiently measured with high accuracy [26] using CB-structured circuits [25], as sketched in Fig. 1. At first, a $+1$ eigenstate of the Pauli operator P_{α} is prepared; then the Clifford layer C to be characterized is applied $d = km$ times (with $k, m \in \mathbb{N}_{>0}$); finally, the expectation value of P_{α} is measured. By implementing PT targeted to both gate and readout errors [8,29,30], thus properly ensuring stochasticity of the latter [31,32], one can directly access the quantity

$$A_{\alpha} \left(\prod_{\beta | P_{\beta} \in \mathcal{R}_{\alpha}^C} f_{\beta}^C \right)^k. \quad (2)$$

Here, the prefactor A_{α} fully accounts for SPAM errors (as constant offsets are eliminated by readout twirling) and $\mathcal{R}_{\alpha}^C$ is the orbit of the operator P_{α} under the application of C [18]. More specifically, we define it as the tuple

$$\mathcal{R}_{\alpha}^C = (P_{\alpha}, \phi(U_C)[P_{\alpha}], \dots, \phi(U_C)^{\circ(m-1)}[P_{\alpha}]), \quad (3)$$

where the orbit's cardinality $m \geq 1$ is the smallest integer such that $\phi(U_C)^{\circ(m)}[P_{\alpha}] = P_{\alpha}$. Note that m is upper bounded by the smallest positive integer u such that $(U_C)^u = I$. By repeating the experiment for different values of k , it is thus

TABLE I. Conjugation of nontrivial Pauli operators in $\mathbb{P}_2$ under a single CZ (described by unitary U_C) and for a single CZ with extra S gates (unitary $\bar{U}_C$). Conjugated Pauli strings in bold are different from the original P_α .

P_α	IX	IY	IZ	XI	XX	XY	XZ	YI	YX	YY	YZ	ZI	ZX	ZY	ZZ
$U_C P_\alpha U_C^\dagger$	ZX	ZY	IZ	XZ	YY	YX	XI	YZ	XY	XX	YI	ZI	IX	IY	ZZ
$\bar{U}_C P_\alpha \bar{U}_C^\dagger$	ZY	ZX	IZ	YZ	XX	XY	YI	XZ	YX	YY	XI	ZI	IY	IX	ZZ

possible to determine with multiplicative precision and in an SPAM-robust way [26,33] the products of the Pauli eigenvalues associated with the operators belonging to each orbit $\mathcal{R}_\alpha^C$. Importantly, if an orbit features at least a pair of Pauli operators with the same support, its cardinality can be reduced by interleaving the layer C with additional layers of specific single-qubit Clifford gates, which can be compiled together with the twirling layers [8,22]. In the remainder of this paper, we refer to this whole procedure as the standard CB protocol for measuring Pauli eigenvalues. The existence of orbits that cannot be reduced to a single element shows that some Pauli eigenvalues cannot be individually learned with high accuracy. This is not a flaw of the standard CB protocol but rather a manifestation of the general limited learnability of Pauli noise, stemming from the existence of a fundamental gauge freedom akin to the one encountered in the context of GST [2,22]. Specifically, Ref. [22] shows that out of the $4^n - 1$ DOF of general Pauli noise, the fundamentally unlearnable ones amount to 2^{n-c} , where c is the number of connected components of the pattern transfer graph based on the orbits analysis. As such, the limited learnability of Pauli noise also impacts alternative protocols developed to measure Pauli eigenvalues. Examples include tensor-network-based noise learning [11] and Averaged Circuit Eigenvalue Sampling (ACES) [9,10,12]. For instance, the ACES protocol can reconstruct all relevant Pauli eigenvalues associated with a local Pauli noise model only by arbitrarily fixing the gauge, e.g., by assuming perfect state preparation [9,12]. A Pauli eigenvalue f_α being unlearnable does not imply that it cannot be estimated. For instance, this can be achieved using a modified CB-structured circuit with $d = 1$, where an eigenstate of P_α is prepared and the expectation value of the conjugated Pauli operator $P_\beta = \phi(U_C)[P_\alpha]$ is measured [8,18,34]. Those alternative approaches, however, lack the SPAM robustness and the multiplicative precision that the standard CB protocol provides for learnable quantities. In what follows, we will collectively refer to strategies used to estimate unlearnable Pauli eigenvalues as low-accuracy protocols, contrasted with the high accuracy of protocols like standard CB.

For the sake of concreteness, we conclude this section by discussing a simple example: the case of a single CZ gate performed on a two-qubit device. In this case, $U_C^2 = I$ limits the cardinality of the orbits to 2, as it emerges also from Table I, which shows the conjugation of all nontrivial Pauli operators in $\mathbb{P}_2$ under a single CZ. Allowing for the usage of additional interleaved S gates, which map $X \leftrightarrow Y$, there are seven Pauli eigenvalues that are directly learnable, i.e., associated with orbits with cardinality 1. The remaining eight eigenvalues are not directly measurable but the standard CB protocols allow to determine, with high accuracy, eight products of eigenvalue pairs. The latter are not all independent from each other since, for example, $f_{IY}f_{ZX} = (f_{IY}f_{ZY})(f_{IX}f_{ZX})/(f_{IX}f_{ZY})$. Still, they

allow to impose a total of six high-accuracy constraints on the parameters of the noise model, leaving us with only two unlearnable DOF, which can be identified, for example, with Pauli eigenvalues f_{IX} and f_{XI} . Those can only be estimated using low-accuracy protocols. More details concerning this simple example are provided in Appendix B.

III. SCALABLE PAULI CHARACTERIZATION

Estimating all the $4^n - 1$ parameters of the Pauli noise map associated with the execution of the layer C on a n -qubit system is not possible at scale. It is therefore necessary to rely on simpler, effective Pauli models featuring a number of parameters that scales more favorably with n . A natural and established way to achieve this is to leverage the locality of the physical noise processes and the resulting sparsity of the model parameters. A prominent example here is the marginalization procedure, a key component of the Cycle Error Reconstruction (CER) scheme [18], which has been successfully used to inform experimental realizations of noise-aware error mitigation technique [23]. Another well-known scalable framework for characterizing Pauli noise is the so-called SPL model, introduced in Ref. [29] and successfully used to perform noise-aware error mitigation on large QPUs [15,35].

A. The SPL model

The limited learnability of Pauli noise also directly impacts effective and scalable models [8,18,35]. Consequently, the implementation of MLCB is beneficial, irrespective of the specific Pauli model under consideration. For the sake of concreteness, however, the remainder of this paper focuses primarily on the SPL model, which we introduce briefly below. The SPL model assumes that Λ_C is generated by a Lindbladian whose jump operators are local and low-weight Pauli operators [8], that is, $\Lambda_C = e^{\mathcal{L}}$ with

$$\mathcal{L}(\rho) = \sum_{\alpha \in \mathcal{K}} \lambda_\alpha^C (P_\alpha \rho P_\alpha - \rho). \quad (4)$$

Here, the set $\mathcal{K}$ consists of all Pauli strings with a support involving up to w neighboring qubits (with respect to the QPU's topology). This also allows the model to capture crosstalk terms. Considering a maximum weight $w = 2$ —a common assumption in the literature [8,15,36] and which we adopt throughout this paper—the number of model parameters associated with Λ_C is given by $|\mathcal{K}| = 3n + 9p$, regardless of the number of gates in the layer, with n the number of qubits and p the number of nearest-neighbor pairs. For a square topology, at scale, we have $p \sim 2n$, which leads to the linear scaling $|\mathcal{K}| \sim 21n$.

The model parameters λ_α cannot be accessed directly but they are related to the Pauli eigenvalues via

$$\log(f_\alpha) = -2 \sum_{\beta \in \mathcal{K}} M_{\alpha\beta} \lambda_\beta, \quad (5)$$

where $M_{\alpha\beta} = \langle \alpha, \beta \rangle_{sp}$ and we omitted the superscript C for brevity. Because of limited learnability, sometimes only products of Pauli eigenvalues are known with high accuracy. It is thus convenient to define the full $|\mathcal{K}|$ -dimensional vector of model parameters $\vec{\lambda}$ [37]. Similarly, we consider an F -dimensional vector $\vec{\xi}$ whose j th element consists of a given product of eigenvalues, associated with the Pauli strings in a given set $\mathcal{X}_j$ (which typically corresponds to some orbit), i.e., $\xi_j = \prod_{\alpha \in \mathcal{X}_j} f_\alpha$. Then, we can generalize Eq. (5) to

$$\log(\vec{\xi}) = -2\vec{M}\vec{\lambda}, \quad (6)$$

where the logarithm is applied elementwise and $\vec{M}$ is a rectangular $F \times |\mathcal{K}|$ matrix whose j th row reads $\vec{M}_{jk} = \sum_{\alpha \in \mathcal{X}_j} M_{\alpha k}$. If enough Pauli eigenvalues (or products thereof) are known such that the rank of $\vec{M}$ is equal to $|\mathcal{K}|$, then Eq. (6) can be inverted and the model parameters determined, typically via a non-negative least square fit

$$\vec{\lambda}(\vec{\xi}) = \operatorname{argmin}_{\lambda_k \geq 0} \|\vec{M}\vec{\lambda} + \log(\vec{\xi})/2\|_2^2. \quad (7)$$

Once all the model parameters $\vec{\lambda}$ are determined, the SPL model proves to be very useful. For example, it can be readily used to effectively amplify or reduce the circuit noise, enabling the execution of error mitigation techniques such as PEC [8,13], tensor-network error mitigation [14] and zero noise extrapolation [15,38] with probabilistic amplification. Indeed, one can show that the (quasi)probabilistic implementation of the map [39]

$$\Lambda_C^{(\beta)}(\rho) = \prod_{k \in \mathcal{K}} (w_k^{(\beta)} \cdot (1 - w_k^{(\beta)}) P_k \cdot P_k) \rho \quad (8)$$

before the layer C effectively scales the overall noise as $\Lambda_C \circ \Lambda_C^{(\beta)} = e^{\beta \mathcal{L}}$, provided that $2w_k^{(\beta)} = 1 - \exp[-2(\beta - 1)\lambda_k]$ [34].

B. Effects of limited learnability on the SPL model

Let us consider the characterization of a Clifford layer C composed of parallel CZ gates—a scenario of significant practical relevance [27,40,41]. In order to fit the corresponding SPL model, it is sufficient to know [8] (1) all the $3n$ eigenvalues associated with weight-one Pauli operators and (2) all products of pairs of $f_\alpha f_\beta$, where $\alpha \in \mathcal{K}$ is a weight-two Pauli string and P_β is the conjugation of P_α under the layer C and, possibly, additional single-qubit Clifford gates [42]. While the standard CB protocol provides us with high-accuracy values for all the products of eigenvalues required in (2), condition (1) is highly impacted by the limited learnability of Pauli noise and some eigenvalues will need to be estimated using alternative, low-accuracy approaches.

In the previous section, we demonstrated that a single, isolated CZ gate has two unlearnable DOF, which can be conveniently associated with the weight-one Pauli eigenvalues f_{XI} and f_{XZ} . Leveraging this fact, along with condition (1), we can immediately compute the number of unlearnable DOF

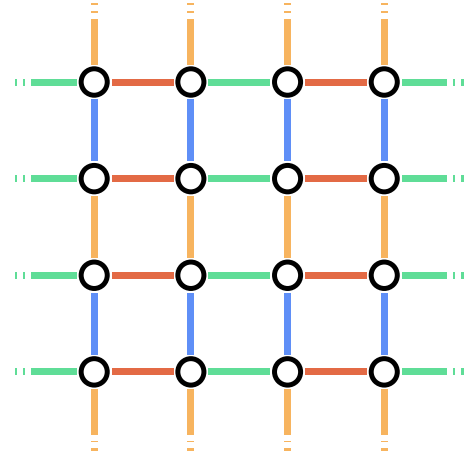


FIG. 2. A QPU with square topology and four layers of parallel CZ gates, identified with blue, green, red, and orange colors, that covers all nearest-neighbor connections between the qubits.

for an SPL characterization of a layer containing n_g parallel CZ gates, which totals $2n_g$. These can also be conveniently associated with $2n_g$ unlearnable weight-one eigenvalues, each one of them associated with a single X operator on one qubit belonging to the layer's support.

Several methods have been proposed to estimate these unlearnable quantities. One approach is the non-SPAM-robust $d = 1$ CB variant mentioned earlier. Another common—but theoretically unjustified—strategy relies on assuming symmetries among the Pauli eigenvalues. For instance, in the case of a single CZ gate, one might assume that unlearnable eigenvalues are equal to their conjugates under the action of C , such as $f_{XI} = f_{XZ}$. This allows the estimation of individual unlearnable eigenvalues via $f_{XI} = \sqrt{f_{XI} f_{XZ}}$, where the product on the right-hand side is learnable. However, since this assumption lacks rigorous justification, the method can introduce significant errors and yields low accuracy.

Regardless of the chosen strategy for addressing limited learnability, once conditions (1) and (2) are satisfied, all information about the (products of) eigenvalues can be encapsulated in the vector $\vec{\xi}$. It is useful to represent $\vec{\xi}$ as a concatenation $\vec{\xi} = (\vec{\xi}, \vec{\zeta})$, where $\vec{\xi}$ contains only learnable (products of) eigenvalues, and $\vec{\zeta}$ comprises unlearnable ones. At this stage, the full-rank matrix $\vec{M}$ can be constructed, enabling the SPL model parameters $\vec{\lambda}$ to be determined by inverting Eq. (6). As shown in Refs. [8,22] and further discussed in Sec. VI, the limited accuracy associated with $\vec{\zeta}$ impacts the overall precision of the scalable error characterization.

So far, we have focused on the characterization of a single layer C . However, in realistic scenarios, multiple distinct layers must often be addressed. A particularly relevant example is depicted in Fig. 2, which illustrates a large n -qubit QPU with a square topology and four layers of parallel CZ gates, color-coded as blue (B), green (G), red (R), and orange (O). These four layers collectively cover all nearest-neighbor connections on the QPU. Combined with arbitrary single-qubit gate layers, they enable the construction of nontrivial circuits that facilitate rapid entanglement distribution across the whole QPU [43]. Such dense configurations are useful for

a variety of applications, including simulating the dynamics of lattice models [15] and implementing variational quantum algorithms [44].

In the limit of large n , the number of CZ gates per layer scales linearly as $n_g \sim 2n$, as does the number of SPL parameters per layer, $|\mathcal{K}| \sim 21n$. Characterizing all four layers requires determining a total of $\sim 84n$ parameters. If each layer is characterized independently, the total number of unlearnable quantities amounts to $2n_g \sim 4n$. Unlike the (non-scalable) full characterization of Pauli noise—where the unlearnable DOF constitute an exponentially small fraction of the learnable ones [22]—this example shows that for large n and considering the SPL model, the ratio of unlearnable DOF to total parameters approaches a constant value of around $4/84 \sim 4.8\%$. As we will discuss in Secs. VI and VII, this small fraction is enough to significantly hinder the overall quality of error characterization and the effectiveness of noise-aware error mitigation techniques.

C. Number of required experiments

Compared to strategies based on depth-one circuits, approaches relying on symmetry assumptions have the significant advantage of not requiring additional experiments. In these cases, the entire experimental runtime is indeed allocated to executing the standard CB protocol. For the sake of concreteness, we still focus on QPUs with a square lattice topology, as illustrated in Fig. 2. In this scenario, it has been shown that the (products of) eigenvalues associated with all orbits $\mathcal{R}_\alpha^C$ for $\alpha \in \mathcal{K}$ can be measured using the standard CB protocol with just eleven CB instances, a single instance referring to given state preparation and measurement stages and potential interleaving of extra single-qubit Clifford gates [8] [45]. For the four layers shown in Fig. 2, this approach requires a total of 44 CB instances.

This number can be further reduced to 16 instances (four per layer) by generalizing the PT procedure to include S gates in the twirling set. This extension symmetrizes the X and Y components of the noise, reducing the amount of information that needs to be extracted and thereby decreasing the experimental runtime [46]. It is important to emphasize that the discussion here focuses solely on the number of unique CB instances required—an essential aspect of implementing the procedure efficiently [46]. However, this does not address the sampling complexity of the characterization strategies, which is also an important factor in determining the overall runtime. A more comprehensive discussion of the sampling complexity associated with Pauli noise characterization can be found in Ref. [47].

IV. MULTILAYER CB

In realistic scenarios, quantum circuits typically involve more than a single (dressed) Clifford layer, requiring the characterization of an entire set $\mathcal{C}$ of distinct layers. The conventional approach to this problem involves characterizing each unique layer individually, following the procedure outlined earlier [8,15]. In this section, we propose augmenting the conventional approach with the MLCB strategy that leverages circuits retaining the fundamental CB structure but relying

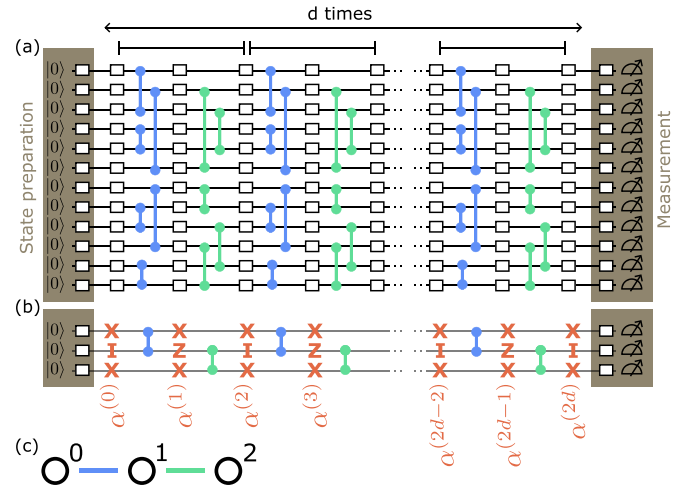


FIG. 3. MLCB protocol: (a) The structure is analogous to the one shown in Fig. 1 with the important difference that, here, the “building block” which is repeated d times is not a single Clifford layer but a combination of more layers (in this specific case two, shown in blue and green). Single-qubit gates used for twirling are indicated in white. (b) Simple scenario where MLCB allows one to measure with high accuracy the product $f_{XIX}^B f_{XZX}^G$ of two eigenvalues, belonging to two different layers, which cannot be accessed with conventional CB protocols. For the sake of simplicity, single-qubit gates used for twirling are now explicitly shown. (c) Simple system consisting of three qubits and two layers.

on building blocks that consist of combinations of different layers, picked from the set $\mathcal{C}$. A simple yet illustrative example is shown in Fig. 3(a), where a combination of two layers, depicted in blue and green, is repeated d times. Note that each layer is independently twirled, as shown by the white boxes indicating single-qubit twirling gates.

To properly describe the MLCB protocol, it is convenient to introduce the definition of a multilayer orbit associated with the pair of layers $L_1, L_2 \in \mathcal{C}$. It generalizes Eq. (3) and reads

$$\mathcal{R}_\alpha^{L_1 L_2} = (P_{\alpha^{(0)}}, P_{\alpha^{(1)}}, \dots, P_{\alpha^{(m-1)}}), \quad (9)$$

where

$$P_{\alpha^{(i)}} = \phi(\bar{U}_{L_1 L_2}^{(j)})[P_\alpha], \quad (10)$$

and the unitary

$$\bar{U}_{L_1 L_2}^{(l)} = \begin{cases} (U_{L_1} U_{L_2})^{l/2}, & \text{if } l \text{ is even,} \\ U_{L_1} (U_{L_2} U_{L_1})^{(l-1)/2}, & \text{if } l \text{ is odd,} \end{cases} \quad (11)$$

captures the effect of two alternating layers $L_1, L_2 \in \mathcal{C}$, associated with unitaries U_{L_1} and U_{L_2} . The cardinality $m \geq 2$ of the multilayer orbit $\mathcal{R}_\alpha^{BG}$ is the smallest even integer such that $\phi(\bar{U}_{L_1 L_2}^{(m)})[P_\alpha] = P_\alpha$. Analogously to the standard CB protocol, by preparing a +1 eigenstate of the operator P_α and measuring its expectation value after the implementation of $m/2$ pairs of twirled layers, the MLCB protocol can thus measure in an SPAM-robust way and with multiplicative precision the

product of the Pauli eigenvalues

$$\mathcal{F}_\alpha^{L_1 L_2} = \prod_{j=0}^{m/2-1} f_{\alpha^{(2j)}}^B f_{\alpha^{(2j+1)}}^G, \quad (12)$$

where the Pauli strings are defined according to Eq. (10).

In the remainder of this section, we demonstrate how this technique can be used to boost the learnability of Pauli noise models by focusing at first on the characterization of a simple system and then discussing its generalization to more complicated and relevant scenarios.

A. Simple example

The simple setup shown in Fig. 3(c) consists of three qubits arranged in an open chain and numbered from 0 to 2. The goal is to characterize the two depicted layers, which correspond to a single CZ between qubits 0 and 1 (blue layer) and between qubits 1 and 2 (green layer). The conventional approach, described in the previous section, prescribes to focus on each layer individually and perform CB protocols to determine with high accuracy all the (products of) eigenvalues related to the orbits that involve the Pauli strings $\alpha \in \mathcal{K}$. Since each layer consists of a single CZ gate, there are four unlearnable DOF, which can be identified with the set of weight-one eigenvalues

$$\{f_{XII}^B, f_{IXI}^B, f_{IXI}^G, f_{IIX}^G\}. \quad (13)$$

Here, we use the notation f_α^L for the Pauli eigenvalues, where α is a Pauli string (ordered according to the increasing qubit indexes) and $L \in \{B, G\}$ indicates the associated layer, i.e., the blue (B) one or the green (G) one. To complete the characterization task, all the eigenvalues in Eq. (13) must be estimated using low-accuracy methods, such as symmetry assumptions or unit-depth variants of CB protocols.

Notably, the MLCB experiment shown in Fig. 3(b) significantly advances the characterization process by enabling the high-accuracy determination of an additional DOF. This is accomplished by analyzing the orbit $\mathcal{R}_{XIX}^{BG} = (P_{XIX}, P_{XZX})$, which yields a precise estimate of the product of individually unlearnable Pauli eigenvalues:

$$\mathcal{F}_{XIX}^{BG} = f_{XIX}^B f_{XZX}^G. \quad (14)$$

These eigenvalues correspond to high-weight, nonlocal Pauli operators, which would be unrelated to the unlearnable DOF in Eq. (13) in the context of a general Pauli noise model. Within the SPL framework, however, not all Pauli eigenvalues are independent from each other and, as we demonstrate in Appendix C, the knowledge of $\mathcal{F}_{XIX}^{BG}$ provided by MLCB is equivalent to the knowing the ratio $\mu_1^{BG} \equiv f_{XII}^B / f_{IIX}^G$ between two unlearnable weight-one eigenvalues in Eq. (13). Specifically, this relationship is expressed as

$$\mu_1^{BG} \equiv \frac{f_{XII}^B}{f_{IIX}^G} = \mathcal{F}_{XIX}^{BG} \frac{(f_{IZI}^G)}{(f_{IIX}^B)(f_{XZI}^G)(f_{IZX}^G f_{IIX}^G)}, \quad (15)$$

where all the terms in round brackets are learnable and can be determined with high accuracy using conventional CB protocols. The nontrivial structure of Eq. (15) arises because the SPL model accounts for crosstalk, which introduces correlated errors also between neighboring qubits 0 and 1. Building

on this insight, we generalize the approach, presenting a systematic method to identify relevant multilayer orbits. This allows MLCB to impose high-accuracy constraints on ratios of otherwise unlearnable weight-one Pauli eigenvalues for arbitrary combinations of Clifford layers.

B. Generalization to two arbitrary layers

We consider the characterization of two generic n -qubit (dressed) layers, identified with blue (B) and green (G) colors. We begin by deriving conclusions applicable to general layers consisting of parallel two-qubit Clifford gates and then proceed by specifically focusing on the relevant case of layers consisting of parallel CZ gates on a QPU with square lattice topology.

A useful way to visualize the problem at hand is to represent each two-qubit gate as a colored edge connecting the two qubits in its support. Those edges collectively form several disconnected graphs G_i over the QPU. An example of this visualization, applied to a square topology, is shown in Fig. 4(a). Since the Clifford layers consist of parallel two-qubit gates, a qubit cannot simultaneously belong to two edges of the same color. Consequently, each graph G_i is a one-dimensional chain, either open or closed, composed of edges with alternating colors. Moreover, closed chains necessarily consist of an even number of qubits. This simple structure, which arises independently of the QPU topology, significantly simplifies the analysis of MLCB protocols, as each graph G_i can be analyzed independently and in parallel. In particular, for each graph G_i we thus need to identify the Pauli string α , and the associated multilayer orbit $\mathcal{R}_\alpha^{BG}$, which allows us to learn DOF that would have been otherwise unlearnable.

In the remainder of this section, we specifically focus on Clifford layers consisting of parallel CZ gates on a square topology. This limits the cardinality of the multilayer orbits to 4, as $(U_B U_G)^2 = I$. In this context, we are able to prove that, for every qubit q in the bulk of a graph G_i , the MLCB protocol enables the high-accuracy determination of one additional DOF that would have been unlearnable using conventional CB methods alone. In particular, this DOF can be identified with the ratio $\mu_q^{(BG)}$ of two individual unlearnable eigenvalues

$$\mu_q^{(BG)} = \frac{f_\beta^B}{f_\gamma^G}. \quad (16)$$

Here, the weight-one Pauli string β (γ) consists only of identities I with the exception of a single X associated with the qubit connected to q via the blue (green) edge. This definition explains the usage of the symbol $\mu_1^{(BG)}$ in Eq. (15), where qubit number 1 is the only one in the bulk of the graph shown in Fig. 3(c).

Remarkably, all of those DOF can be accessed by preparing Pauli strings α consisting solely of I and X operators, regardless of the specific qubit q or graph G_i . Consequently, by preparing a $+1$ eigenstate of $X^{\otimes n}$ and measuring in the X basis, it becomes possible to determine all the eigenvalue ratios $\mu_q^{(BG)}$ simultaneously, with a single MLCB execution. This parallelization introduces only a limited overhead compared to the conventional procedure. Together with the details in Appendix C, the remainder of this section provides a

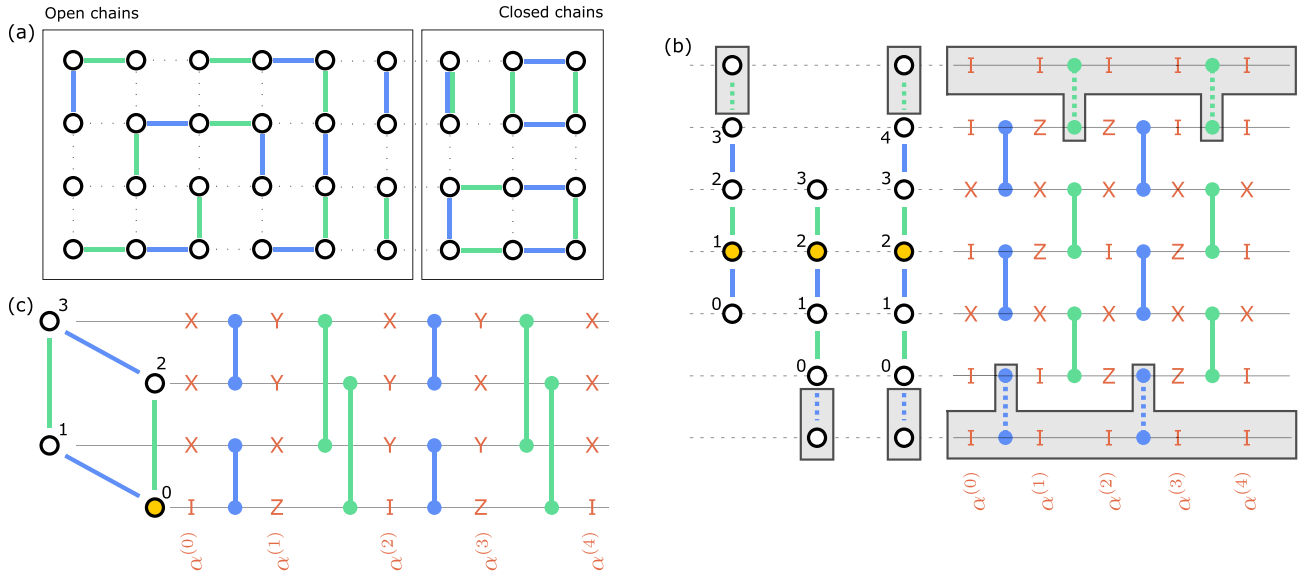


FIG. 4. (a) An example of several disconnected graphs G_i obtained by considering edges associated with two layers, blue and green, of parallel two-qubit gates. The graphs can be divided into open and closed chains, as shown by the black rectangles. (b) Analysis of open chains showing the conjugations of specific Pauli strings $\alpha^{(m)}$ (with $m = 0, \dots, 4$) under four subsequent layers, in the pattern BGBG, that constitute the building block of MLCB protocols. Different chains, featuring four and five qubits, are depicted on the left (each with a specific qubit numbering). By looking at the corresponding qubits (and gates) on the right side, one can identify the circuit blocks that allow to measure specific products of eigenvalues. The latter can then be used to determine the ratios μ_q^{BG} of unlearnable eigenvalues associated with the qubit q (highlighted in yellow). Gray-shaded areas show scenarios in which the presence of additional qubit is trivial. (c) Analysis of closed chains consisting of four qubits. As for the previous panel, the circuits show how MLCB can measure a product of four eigenvalues, which can then be used to determine the ratio μ_q^{BG} (with $q = 0$ highlighted in yellow).

concrete and operational framework for implementing MLCB in relevant scenarios, highlighting the protocol's capability to unlock additional DOF in an efficient way.

1. Open chains

Let us start with the analysis of open chains, some of which are depicted in the left rectangle of Fig. 4(a). Open chains consisting of just two qubits feature a single layer (and have no bulk), meaning that MLCB cannot be implemented. As for open chains with three qubits, they have already been fully covered in Sec. IV A. We therefore proceed by focusing on open chains featuring a total of four qubits. To analyze them within the MLCB framework, it is necessary to repeatedly implement the minimal four-layer block with a BGBG pattern, as shown in Fig. 4(b). If the open chain features two blue CZ gates, Fig. 4(b) shows how MLCB can directly measure the product of four eigenvalues $\alpha_4 = f_{IXIX}^B f_{XZXZ}^G f_{XIXZ}^B f_{XZXI}^G$. In Appendix C, we prove it to be equivalent to the ratio $\mu_1^{(BG)}$ between two unlearnable eigenvalues associated with the bulk qubit 1. Similarly, if the open chain features two green CZ gates, MLCB can directly measure $\alpha'_4 = f_{IXIX}^B f_{IXZX}^G f_{ZXIX}^B f_{ZXZX}^G$, which we prove to be equivalent to the ratio $\mu_2^{(BG)}$. By exploiting inversion symmetry, analogous analyses can be performed to determine the ratios of unlearnable eigenvalues associated with the other qubits in the bulk of four-qubit chains.

As for open chains consisting of five qubits, as shown in Fig. 4(b), there are three qubits in the bulk. The ratios $\mu_q^{(BG)}$ associated with qubits $q = 1, 3$ can be readily determined by following the strategies discussed above for the four-qubit

chains, as the relevant Pauli strings $\alpha^{(m)}$ always have an identity I operator on the qubit at the opposite side of the chain (see the gray-shaded areas). In contrast, to determine the ratio associated with the central qubit, all the five qubits are involved. In particular, MLCB can directly measure the product $\alpha_5 = f_{IXIX}^B f_{IXZX}^G f_{ZXIX}^B f_{ZXZX}^G$, which we then prove in the Appendix C to be equivalent to $\mu_2^{(BG)}$. Interestingly, even if more qubits were present (see the gray-shaded areas), the relevant Pauli strings $\alpha^{(m)}$ always feature operators different from the identity I on no more than five qubits. As a result, when analyzing longer open chains, we can readily determine the ratio $\mu_q^{(BG)}$ for each qubit in the bulk by resorting to the MLCB strategies detailed so far.

2. Closed chains

The shortest closed chain consists of just two qubits, both residing in the bulk, each paired by a CZ gate in both layers [see in the upper left corner of the “Closed chains” rectangle in Fig. 4(a)]. In this straightforward case, the MLCB approach directly provides us with two products of eigenvalues, $c_2 = f_{IX}^B f_{ZX}^G$ and $c'_2 = f_{XI}^B f_{XZ}^G$, which are equivalent to the two ratios $\mu_0^{(BG)}$ and $\mu_1^{(BG)}$. The analysis of chains consisting of four qubits is more involved. In the case shown in Fig. 4(c), MLCB allows one to measure $c_4 = f_{IXXX}^B f_{ZXYY}^G f_{IYYX}^B f_{ZYXY}^G$, a product of eigenvalues that is equivalent to the ratio $\mu_q^{(BG)}$ of unlearnable eigenvalues associated with qubit $q = 0$. By rotating and/or mirroring the chain, one can systematically derive the ratios for the other three qubits q in the bulk of the chain. Detailed derivation is provided in Appendix C. For longer chains involving six or more qubits, the results derived

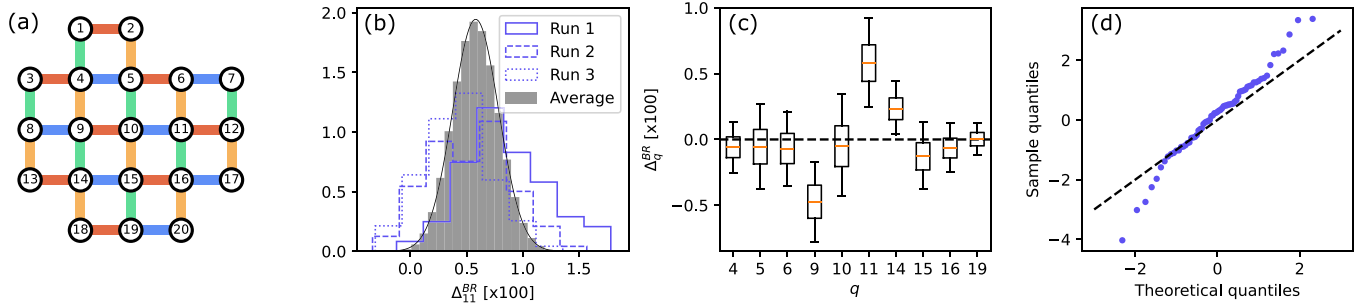


FIG. 5. (a) Layout of the IQM Garnet™ [27] quantum processor, highlighting the four characterized layers: blue (B), green (G), red (R), and orange (O). (b) Bootstrapped distributions of the difference Δ_{11}^{BR} between products of high-weight Pauli eigenvalues, as measured by MLCB, for three consecutive experimental runs (purple) and their average (gray). (c) All measured differences Δ_q^{BR} for the BR layer pair. (d) Q - Q plot of the ratios $\Omega_q^{L_1 L_2}$ [defined in Eq. (18)] for all layer pairs {BR, BO, BG, RO, RG, OG} and a reference standard normal distribution (black dashed line).

for open chains can be directly applied. This is because the latter never involves more than five qubits at a time, regardless of the chain length. A clear example of this can be seen from Fig. 4(b) by identifying (as a single-qubit number 5) the two gray-shaded qubits, effectively showing how to use MLCB to determine the ratio $\mu_q^{(CB)}$ associated with each qubit in a bulk of a six-qubit closed chain.

C. More than two layers

The approach detailed so far can be readily generalized to scenarios featuring more than two layers. Let us focus on a specific qubit q , which belongs to the support of l layers in the set $\{L_1, \dots, L_l\}$. By simply implementing the procedure outlined before, for each pair of layers $(L_i L_j)$ we can thus determine the ratio $\mu_q^{(L_i L_j)}$, which would have been unlearnable if only standard CB were used. Unfortunately, not all ratios $\mu_q^{(L_i L_j)}$ are independent of each other. Specifically, the knowledge of the $l-1$ ratios $\mu_q^{(L_i L_{i+1})}$ (for $i = 1, \dots, l-1$) is enough to reconstruct all the remaining ones, meaning that MLCB provides us with $l-1$ DOF that would have been unlearnable otherwise.

Let us discuss the concrete effect of MLCB on the highly structured example in Fig. 2(a), where every qubit q in the bulk of the QPU belongs to the support of each one of the four layers, blue (B), green (G), red (R), and orange (O). By performing MLCB on three pairs of layers, we can determine three independent ratios per qubit, say $\mu_q^{(BG)}$, $\mu_q^{(GR)}$, and $\mu_q^{(RO)}$. In the limit of large n , this brings down the total number of unlearnable DOF that needs to be determined with low-accuracy methods from $\sim 4n$ to $\sim n$. Importantly, this 75% reduction of unlearnable DOF comes with a rather small experimental overhead, around 7% more circuit runs, since it requires to perform only three extra MLCB experiments on top of the 44 CB instances required in any case by the standard approach. This shows how MLCB can be beneficial, especially in the highly relevant context of characterizing errors during the execution of circuits with few dense layers of CZ gates.

V. EXPERIMENTAL IMPLEMENTATION

We experimentally validate the MLCB protocol on IQM Garnet™, a 20-qubit superconducting quantum computer

[27]. Our objectives are twofold: first, to use MLCB to obtain high-accuracy data inaccessible to conventional CB protocols; and second, to test the common—but theoretically unjustified—assumption of symmetry between unlearnable Pauli eigenvalues. The protocol is robust against both temporal drifts in system noise and deviations from the underlying effective Pauli noise model. We focus on the four gate layers depicted in Fig. 5(a), which pairwise define open chains of various lengths. Under the SPL model, for each qubit q in the bulk of a chain formed by layers L_1 and L_2 , MLCB enables us to learn the ratio $\mu_q^{(L_1 L_2)}$ of weight-one eigenvalues, which are unlearnable using conventional CB. This is achieved by directly measuring products of nonlocal and high-weight Pauli eigenvalues $\mathcal{F}_{\alpha(q)}^{L_1 L_2}$, without assuming a specific Pauli noise model. Importantly, the state-preparation and measurement stages can be chosen such that the same experimental MLCB data also yield the product $\mathcal{F}_{\beta(q)}^{L_1 L_2}$, where the initial Pauli operators are related by conjugation, $P_{\beta(q)} = \phi(U_{L_1})[P_{\alpha(q)}]$. Conventional CB protocols cannot resolve these quantities individually; they can only measure the overall product $f = \mathcal{F}_{\alpha(q)}^{L_1 L_2} \mathcal{F}_{\beta(q)}^{L_1 L_2}$. Without MLCB, estimating the individual terms requires low-accuracy methods, such as assuming $\mathcal{F}_{\alpha(q)}^{L_1 L_2} \approx \mathcal{F}_{\beta(q)}^{L_1 L_2} \approx \sqrt{f}$. We directly test this assumption by defining their difference as

$$\Delta_q^{L_1 L_2} = \mathcal{F}_{\alpha(q)}^{L_1 L_2} - \mathcal{F}_{\beta(q)}^{L_1 L_2}. \quad (17)$$

A measured value of $\Delta_q^{L_1 L_2}$ statistically incompatible with zero provides direct, Pauli model-independent, and time-drift-robust evidence that the symmetry assumption is violated and that MLCB yields information beyond the reach of conventional methods. This approach is inherently robust against temporal variations, as differences are computed from the same dataset, eliminating potential systematic errors from comparing distinct datasets obtained under different noise conditions. For each pair of layers in Fig. 5(a), we perform three back-to-back MLCB experiments. While each measured difference $\Delta_q^{L_1 L_2}$ is inherently drift-robust, this approach also allows us to probe the temporal stability of $\Delta_q^{L_1 L_2}$ while accumulating additional statistics. Figure 5(b) exemplifies this for Δ_{11}^{BR} , associated with qubit $q = 11$ and the blue (B) and red (R) layers. After confirming mutual consistency among the differences from each run, we average the data to

improve statistical precision. The results of each experiment are bootstrapped 100 times with respect to both shots and random Pauli twirling instances, yielding the three purple histograms in Fig. 5(b) (one per run). The data from the three runs are then averaged, resulting in the gray histogram, which is well described by a Gaussian (fitted black line). We compute the ratio

$$\Omega_q^{L_1 L_2} = \text{Mean}(\Delta_q^{L_1 L_2}) / \text{Std}(\Delta_q^{L_1 L_2}), \quad (18)$$

which for Δ_{11}^{BR} is $\Omega_{11}^{BR} \sim 2.84$, corresponding to a probability of Δ_{11}^{BR} not being compatible with zero of approximately 99.5%. As detailed in Appendix D, we extend this analysis to all measured $\Delta_q^{L_1 L_2}$. The box plot in Fig. 5(c) summarizes the results for the BR layer pair, with whiskers indicating the 95% confidence interval derived from bootstrapping, showing another difference, Δ_9^{BR} , which is most likely not compatible with zero. We verify that the bootstrapped distributions are Gaussian (with the exception of one manifestly non-Gaussian distribution, Δ_5^{OR} , which we then exclude from the following analysis). The Q - Q plot in Fig. 5(d) shows the distribution of all the ratios $\Omega_q^{L_1 L_2}$. With 67 ratios, the theoretical quantiles on the x axis (computed assuming a normal distribution and Filliben's estimate) lie in the interval $(-2.32, 2.32)$, while the experimental $\Omega_q^{L_1 L_2}$ on the y axis span a much larger range $(-4.03, 3.39)$. This confirms the existence of several differences statistically incompatible with zero, demonstrating that MLCB robustly measures asymmetries between Pauli eigenvalues inaccessible to conventional characterization. Our experimental results (detailed in Appendix D) show that the products of Pauli eigenvalues measured by MLCB on IQM GarnetTM deviate from unity by approximately 0.1. However, the observed differences $\Delta_q^{L_1 L_2}$ are one-to-two orders of magnitude smaller, consistently below 0.006. This suggests that while MLCB provides a statistically significant improvement in accuracy over conventional methods, the full extent of this gain may be partially obscured by other experimental nonidealities, such as temporal instabilities and deviations from the considered effective noise model.

VI. IMPROVED ERROR CHARACTERIZATION

Having theoretically established that MLCB can significantly reduce the number of unlearnable DOF in realistic scenarios and having demonstrated that these protocols can be implemented in practice, we now investigate to what extent this additional knowledge improves the overall accuracy of the fitted SPL model. This is not a trivial step, as it requires combining high-accuracy data from CB and MLCB with lower-accuracy methods needed to compensate for the remaining unlearnable components of the Pauli noise.

We address this point through numerical simulations, which provide a controlled environment free from other experimental nonidealities and where the exact noise model is known for reference—a condition not met in a real device. To this end, we numerically simulate the characterization of a 20-qubit IQM GarnetTM QPU [27]. In particular, we consider two different dense layer structures, depicted in the insets of Fig. 6(a). The leftmost configuration corresponds to the one used for the experiment in Sec. V, with pairs of layers defining open chains, whose MLCB analysis follows the scheme in

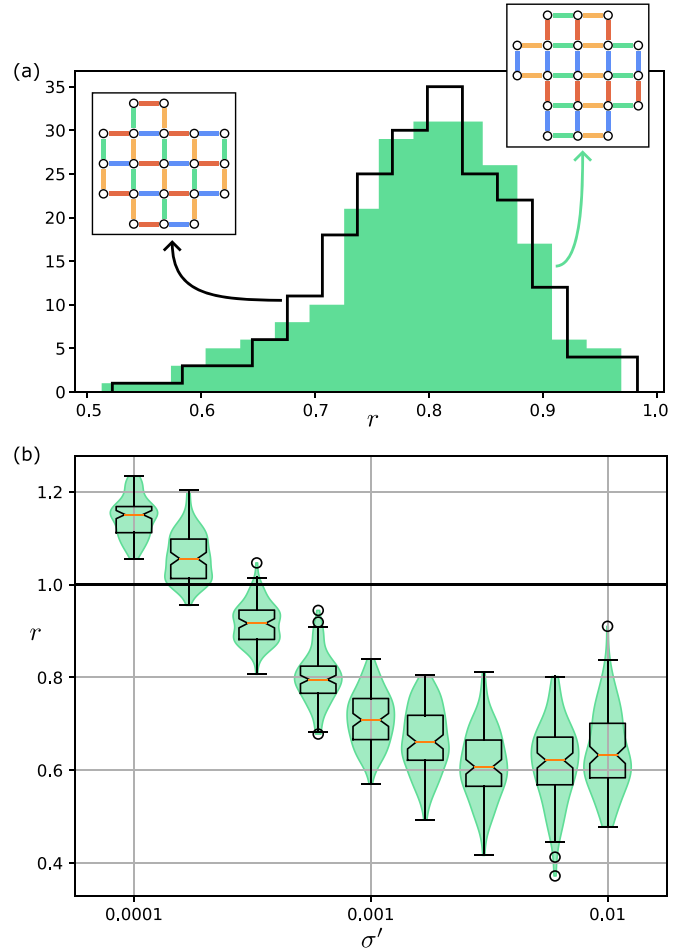


FIG. 6. (a) Reduction of the L_1 distance between the exact and characterized noise model enabled by MLCB, with respect to the conventional CB approach and symmetry assumption. We consider 200 randomly generated noise models and plot the ratio $r = \Delta_m / \Delta_c$ [see Eq. (19)] for each of the two-layer configurations shown in the insets. (b) Similarly to the previous panel, we plot ratios r associated with conventional strategies that allow measuring unlearnable quantities with accuracy $\sigma' \geq \sigma = 10^{-4}$ (e.g., via unit-depth circuits), for the configuration shown in the left inset of panel (a).

Fig. 4(b). In contrast, the rightmost configuration features closed chains, which are analyzed according to the scheme in Fig. 4(c).

Our numerical simulations consist of five steps. At first, (1) we randomly generate a realistic SPL noise model; i.e., we associate a map Λ_L to each one of the four layer $L \in \{B, G, R, O\}$ based on model parameters $\tilde{\lambda}^L$ that mimics the behavior of realistic devices (see Appendix F for more details). (2) From each exact noise map Λ_L we compute all the (products of) eigenvalues that are needed to fit the SPL model. In an actual experiment, those would be obtained using conventional CB, MLCB, as well as possible alternative low-accuracy methods. (3) We artificially add uncertainties independently to all (product of) eigenvalues, in the form of Gaussian noise, to simulate the statistical uncertainty associated with an actual experimental characterization. For (products of) eigenvalues determined via conventional CB

and MLCB, we consider a fixed standard deviation σ . To simulate the implementation of low-accuracy strategies that involve performing extra measurements, like the one based on unit-depth CB-like circuits, we account for the additional uncertainties due to SPAM noise and the lack of multiplicative precision by considering a larger standard deviation $\sigma' \geq \sigma$. (4) We then use these noisy eigenvalues to fit two SPL models for each layer: Λ_L^c , which is based on parameters $\{\tilde{\lambda}_c^L\}$ fitted using solely conventional CB data and low-accuracy strategies, and Λ_L^m , whose fitted parameters $\{\tilde{\lambda}_m^L\}$ also leverages the additional knowledge associated with MLCB experiments. (5) Finally, we compute the L_1 distances Δ_t , with $t = c, m$, between the exact and the reconstructed noise models, summed over all the layers $L \in \{B, G, R, O\}$ as

$$\Delta_t = \sum_L \|\tilde{\lambda}^L - \tilde{\lambda}_t^L\|_1. \quad (19)$$

We also define their ratio as $r = \Delta_m / \Delta_c$.

The most intricate aspect of this otherwise straightforward procedure lies in step (4), particularly the role of products of eigenvalues measured with MLCB during the model fitting process. Several strategies can be adopted to address this step. One straightforward method is to treat experimental data from conventional CB, MLCB, and low-accuracy methods on the same ground, by generalizing Eq. (6) to the multilayer scenario

$$\log(\tilde{\xi}^+) = -2\tilde{M}^+ \tilde{\lambda}^+. \quad (20)$$

Here, $\tilde{\lambda}^+$ is the concatenation of the parameter vectors $\tilde{\lambda}_L$, associated with each individual layer $L \in \{B, G, R, O\}$. Similarly, $\tilde{\xi}^+$ is the concatenation of the (products of) eigenvalues vectors $\tilde{\xi}_L$, which are used in the conventional approach, supplemented with the products of eigenvalues $\mathcal{F}_\alpha^{L_1 L_2}$ measured using MLCB. Eq. (20) can then be solved using non-negative least square fitting in a $4|\mathcal{K}|$ -dimensional space.

Alternatively, for each layer L , one can first identify a specific set of unlearnable eigenvalues $\{f_\beta^L\}$ required to fit the SPL model. For example, as discussed above, one can focus on all weight-one Pauli strings β consisting of a single X operator on the layer's support. Since these eigenvalues are inherently unlearnable, their initial estimates rely on low-accuracy data. However, they can be significantly improved by imposing the high-accuracy constraints provided by MLCB, e.g., $\mu_q^{(L_1 L_2)}$. This optimization step, detailed in Appendix E, yields refined eigenvalue estimates $\{f_\beta^L\}$, which can then be used for fitting the SPL model layer by layer. In the remainder of this work, we primarily adopt this second approach, which places greater emphasis on leveraging high-accuracy MLCB data over low-accuracy estimates. However, other fitting strategies are certainly possible. A thorough exploration of such strategies represents a compelling avenue for future research, as it could unlock further improvements in model fitting accuracy and performance.

The numerical results are summarized in Fig. 6, where we plot the values of the ratio r for different scenarios. In panel (a), we visualize the values of r for 200 randomly generated noise model, considering both the layer structures shown in the two insets. We consider a reasonable experimental uncertainty of $\sigma = 10^{-4}$ [8] and use the symmetry assumption to

estimate the unlearnable individual eigenvalues. Regardless of the specific layer structure, we observe a r consistently below 1, showing that the possibility to measure with high-accuracy extra DOF, unlocked by MLCB, leads to a better overall error characterization. In Fig. 6(b), we simulate a characterization procedure that relies on extra measurements, for example, via unit-depth CB-like experiments, to estimate with low accuracy the required unlearnable eigenvalues. For a fixed $\sigma = 10^{-4}$, we consider different uncertainties associated with unlearnable quantities $\sigma \leq \sigma' \leq 100\sigma$ and compute the ratio r for 100 randomly generated noise models for each value of σ' . Provided that $\sigma' \gtrsim 3\sigma$, which makes the distinction between high- and low-accuracy procedures meaningful and justified, we observe values of the ratio r well below 1, which saturates to a range $0.5 \lesssim 0.8$ in the realistic regime $\sigma' \gtrsim 10\sigma$. This shows again how MLCB can significantly improve the error characterization. In contrast, in the extreme and unrealistic regime $\sigma \sim \sigma'$, MLCB can even be disadvantageous due to the amplification of the σ error associated with the computation of ratios $\mu_1^{(L_1, L_2)}$ (see Appendix C), which makes it more convenient to just consider the direct measurements of the unlearnable eigenvalues.

VII. EFFECTS ON ERROR MITIGATION

Error characterization plays a crucial role in informing noise-aware error mitigation strategies, prompting an important question: how does MLCB and the associated improved characterization impact error mitigation performance? To address this, we conducted a series of numerical experiments analyzing the performance of PEC [8,13,15,23] on Clifford circuits. The choice of Clifford circuits is particularly advantageous because evaluating PEC performance in this context is both conceptually straightforward and computationally efficient [28]. Moreover, generic non-Clifford circuits can always be decomposed into a sum of maps associated with Clifford circuits, making the following analysis relevant, at least at a qualitative level, also in practical and general use cases. More precisely, data obtained on Clifford circuits can be indicative of worst-case scenario performances [28].

In the analysis, we focus on circuits consisting of the four (dressed) Clifford layers depicted in the rightmost inset of Fig. 6(a), each one associated with a randomly generated noise map Λ_L . As discussed in the previous section, for each layer $L \in \{B, G, Y, R\}$ we also define a low-accuracy reconstructed noise map Λ_L^c as well as another reconstructed map Λ_L^m that makes use of the extra high-accuracy DOF provided by MLCB. When using PEC to mitigate errors on a given layer, we probabilistically modify the circuit such that the noisy execution of the layer is effectively composed with the inverse of the characterized noise map Λ_L^t , with $t \in \{c, m\}$. As a result, the execution of each layer is effectively associated with the map $\xi_L^{(t)} = \Phi(U_L) \circ \Lambda_L \circ (\Lambda_L^t)^{-1}$, which corresponds to the noiseless layer execution in the limit of perfect characterization, i.e., $\Lambda_L^t = \Lambda_L$. The effect of such a map on a Pauli operator P_α simply reads

$$\xi_L^{(t)}[P_\alpha] = P_\alpha R_{\alpha,t}^L, \quad (21)$$

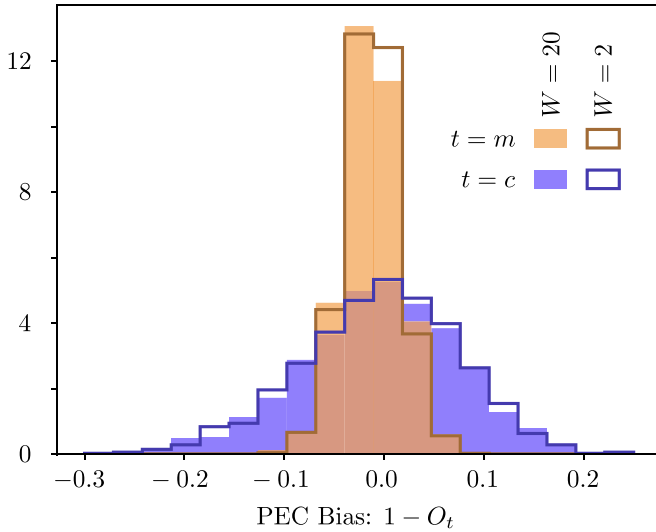


FIG. 7. Bias $1 - O_t$ resulting from the application of PEC on random Clifford circuits under 200 random SPL noise models, comparing two characterization methods: the conventional procedure ($t = c$) and the advanced one leveraging the additional learnability provided by MLCB ($t = m$). The MLCB-based method achieves improved error mitigation accuracy, consistently outperforming the standard procedure regardless of the observable's weight, whether low ($W = 2$) or high ($W = 20$).

where $P_{\alpha'} = \Phi(U_c)[P_{\alpha}]$ is the conjugation of P_{α} under the Clifford unitary and

$$R_{\alpha,t}^L = \frac{f_{\alpha}^L}{f_{\alpha,t}^L} \quad (22)$$

is the ratio between the exact Pauli eigenvalue f_{α}^L , derived from Λ_L , and the eigenvalue $f_{\alpha,t}^L$ obtained from the characterized model $\Lambda_{L,t}^t$, with $t \in \{c, m\}$.

We consider Clifford circuits, consisting of the repeated implementations of J Clifford layers $\bar{L}_j$, with $j \in [1, J]$. Each $\bar{L}_i$ is the concatenation of a layer randomly sampled from $\{B, G, R, O\}$ with a random layer consisting of single-qubit Clifford gates. By preparing a $+1$ eigenstate of a Pauli operator $P_{\beta^{(0)}}$ and measuring the expectation value O of its conjugation $P_{\beta^{(t)}}$ with respect to all Clifford layers, we obtain (neglecting SPAM errors and applying PEC)

$$O_t = \prod_{i=0}^{J-1} R_{\beta^{(i)},t}^L. \quad (23)$$

Here, $\beta^{(i)}$ is the Pauli string obtained under conjugation by the first $i \geq 0$ Clifford layers, i.e., $P_{\beta^{(i)}} = \Phi(\prod_{j=1}^i U_{L_j})[P_{\beta^{(0)}}]$. A perfect characterization would result in the noiseless result $O = 1$ and deviations from this value are a consequence of the finite accuracy of the reconstructed maps $\Lambda_{L,t}^t$.

By computing O_c and O_m for different noise models and different circuits, we observed that the latter features a much smaller standard deviation, indicating that the extra learnability offered by MLCB can directly result in a better performance of PEC. The results of this numerical analysis are shown in Fig. 7, where we consider the same 200 randomly generated noise models employed for Fig. 6(a). For each

noise model, we calculated O_t across ten random circuits, with each circuit composed of 40 random Clifford layers $\bar{L}_j$. Additionally, each circuit began with a randomly chosen initial Pauli string $\beta^{(0)}$, subject only to the condition that the final measured Pauli string $\beta^{(J)}$ had a specified weight W . The latter parameter does not seem to play a significant role. The standard deviation associated with O_c is $\text{std}(O_c) = 0.08$ for both $W = 2$ and $W = 20$. In contrast, the standard deviation associated with O_m is $\text{std}(O_m) = 0.03$ for both $W = 2$ and $W = 20$, showing almost a threefold improvement over the conventional scenario.

VIII. DISCUSSION

In this paper, we have introduced the MLCB protocol and showed its potential in increasing the accuracy of characterization of effective Pauli noise models, thanks to improved learnability. Numerical simulations and a simple experiment have been performed focusing on practically relevant scenarios, featuring dense Clifford layers consisting of parallel CZ gates implemented on a QPU with square topology. In these cases, we show in detail how to set up the MLCB procedure to significantly decrease the number of unlearnable DOF while keeping the runtime overhead at a minimum. Our results showcase a scalable and resource-efficient procedure that can be readily extended to various combinations of CZ layers with minimal overhead.

While our focus has been on CZ gates and (SPL) models with weight-two generators, MLCB can be generalized to other Clifford gates, higher-weight SPL models or alternative effective Pauli noise frameworks like CER [18]. These extensions, though valuable, are beyond the scope of this paper. On the other hand, MLCB is naturally compatible with generalizations of PT beyond the Pauli group with which the independent number of noise parameters can be reduced [46], as exemplified in Sec. III A for CZ gates and considering the addition of S gates to the twirling set.

It is also important to recognize that limited learnability is not the sole factor affecting characterization accuracy and the performance of noise-aware error mitigation: Time-dependent parameter drift, as observed in our experiment, and out-of-model noise contributions can also play a significant role. Recent work, such as Ref. [28], has explored these two issues in the context of PEC. In this sense, our discussion in Sec. VI complements these analyses by focusing on the effect of limited learnability. In general, it is important to stress that practical implementations must consider all these factors holistically to ensure robust characterization and effective noise mitigation.

The MLCB implementation presented in this work is both practical and resource efficient, but there is room for optimization in several areas.

First, we have not addressed the optimal allocation of experimental resources, such as the distribution of shots and rounds of randomized compiling (RC), between different CB and MLCB protocols. In this regard, insights from the statistical analyses in protocols like ACES, as detailed in Ref. [12], provide valuable directions for future work. Clearly, optimizing the resource allocation could enhance the precision of characterization within a finite experimental runtime. The

implementation of PT at the hardware level, reaching the ideal limit of a single shot per randomized circuit, is also likely to significantly boost the precision of the characterization protocol with a fixed runtime [48,49]. More broadly, the study of sampling complexity associated with quantum noise characterization is an active and compelling area of research, with recent advancements detailed, e.g., in Refs. [47,50,51].

Second, as discussed in Sec. VI, efficiently fitting noise model parameters based on measured (products of) eigenvalues is a nontrivial task. While we propose a weighting strategy that prioritizes high-accuracy data (obtained through CB and MLCB) over lower-accuracy data from other methods, this approach, validated through numerical simulations of random noise models, is not guaranteed to be optimal. Further exploration in this area could yield more robust and efficient fitting techniques, enhancing the overall utility of MLCB.

Finally, while our implementation significantly reduces the number of unlearnable eigenvalues with only a minor runtime increase, we do not exclude the possibility of alternative strategies that could unlock even more DOF, potentially at the cost of increased experimental runtime. In this respect, during the finalization of this work, we became aware of theoretical advancements in Ref. [52], which propose a general framework for computing the learnability of Pauli noise models, considering also the SPAM parameters as target of learning. Our results are consistent with their findings in overlapping cases, such as the characterization of quasilocal noise on n -qubit systems arranged in 1D rings with two dense layers of parallel CZ gates, thus suggesting the optimality of MLCB in this scenario. In general, we believe our work complements the theoretical framework in Ref. [52] by providing practical details for characterizing general layers on 2D QPUs with square topologies, thus enabling straightforward implementation on physical devices, and by quantifying the impact of MLCB's enhanced learnability on characterization quality and its effectiveness in error mitigation within the widely adopted SPL model.

ACKNOWLEDGMENTS

A.C. gratefully acknowledges S. Chen and A. Kandala for their insightful discussions during the APS March Meeting 2024, where the foundational ideas of this work were presented. The authors acknowledge support from the German Federal Ministry of Education and Research (BMBF) under Q-Exa (Grant No. 13N16062) and QSolid (Grant No. 13N16161). The authors also acknowledge the entire IQM Technology team for their support in the development of this work.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain commercially sensitive information. The data are available from the authors upon reasonable request.

APPENDIX A: PAULI TWIRLING

Here, we briefly detail how the noise associated with the execution Clifford layer C can be recast into a Pauli

channel via twirling. Formally, the twirling of the map $\phi(U_C)$ is performed by sandwiching C in between two layers of single-qubit operators randomly sampled from the twirling set $\mathbb{T}$. Unless explicitly stated otherwise, in this work we focus on the standard PT, with $\mathbb{T}$ consisting of all possible product combinations of single-qubit Pauli operators, i.e., $\mathbb{T} = \mathbb{P}_n := \{I, X, Y, Z\}^{\otimes n}$. Instead of executing U_C , we first implement a twirling layer $T \in \mathbb{T}$, followed by U_C and then by the conjugated layer $T_C := U_C T U_C^\dagger \in \mathbb{T}$, which also belongs to $\mathbb{T}$ since the Pauli group is invariant under Clifford conjugation. If we assume that the noise map Ω associated with the execution of a layer T of single-qubit gates is gate independent, i.e., $\nu(T) := \Omega \circ \phi(T) \forall T$, one can show that uniformly sampling T from the twirling set $\mathbb{T}$ and subsequently averaging the results, denoted as $\langle \cdot \rangle_{T \in \mathbb{T}}$, leads to the following effective evolution:

$$\begin{aligned} & \langle \nu(T_C) \circ \mu(U_C) \circ \nu(T) \rangle_{T \in \mathbb{T}} \\ &= \Omega \circ \phi(U_C) \circ \langle \phi(T) \circ \tilde{\Lambda}_C \circ \Omega \circ \phi(T) \rangle_{T \in \mathbb{T}} \\ &= \Omega \circ \phi(U_C) \circ \Lambda_C, \end{aligned} \quad (\text{A1})$$

where

$$\Lambda_C[\rho] = \sum_{\alpha} p_{\alpha} P_{\alpha} \rho P_{\alpha}^{\dagger} \quad (P_{\alpha} \in \mathbb{P}_n) \quad (\text{A2})$$

is an effective Pauli channel associated with the execution of the Pauli-dressed Clifford layer C . The channel Λ_C corresponds to the noise of the original Clifford layer C , which contributes the majority of the noise, together with the noise of the twirling layer T .

APPENDIX B: THE EXAMPLE OF A SINGLE CZ GATE

Let us provide more detail about the simple-yet-instructive example of a system with $n = 2$ qubits and on a layer C consisting of a single CZ gate between the two qubits. The 15 nontrivial Pauli operators $P_{\alpha} \in \mathbb{P}_2$ are listed in the first row of Table I, while the second row shows their conjugation under C . Since there are only three orbits with cardinality one, associated with the Pauli strings $\alpha \in \{IZ, ZI, ZZ\}$, the three corresponding Pauli eigenvalues are the only ones that can be determined with high accuracy using standard CB directly on the layer C . See, for example, the setup in Fig. 8(a) that can measure $A_{ZZ}(f_{ZZ})^d$ and, thus, determine with high accuracy the individual eigenvalue f_{ZZ} . All other orbits have cardinality 2 (which is the largest possible since, in this case, $U_C^2 = I$). In several cases, however, the Pauli operators belonging to the same orbit share the same two-qubit support (for example, $\mathcal{R}_{XX}^C = \mathcal{R}_{YY}^C = \{P_{XX}, P_{YY}\}$). In these cases, the orbit's cardinality can be reduced to one by inserting two extra S gates after each execution of the layer C (and then recompiling them with the following layer of single-qubit gates). This modifies the associated unitary as $U_{\tilde{C}} = (S \otimes S)U_C$, providing us with the third row of Table I and, importantly, with four orbits with size one: $\mathcal{R}_{\alpha}^{\tilde{C}}$ for $\alpha \in \{XX, XY, YX, YY\}$. An example of an interleaved CB circuit that allows determining f_{XX} is shown in Fig. 8(b). At this point, we are left with eight Pauli strings, which can be conveniently organized into two sets $\mathcal{S}_A = \{IX, IY, ZX, ZY\}$ and $\mathcal{S}_B = \{XI, YI, XZ, YZ\}$, whose associated orbits cannot be reduced to size one. This means

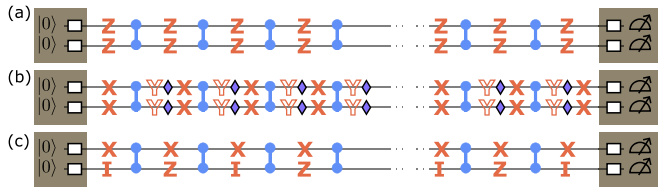


FIG. 8. (a) Simple example of a CB circuit that allows for the high-accuracy determination of Pauli eigenvalue f_{ZZ} . For the sake of clarity, here we do not display the white boxes associated with RC and, instead, we show (in red) the Pauli operator that is measured at the end and how it propagates back (considering its conjugation under the CZ gates) all the way to the state preparation stage. (b) Similarly to the previous panel, we show how interleaved CB can determine the eigenvalue f_{XX} with high accuracy. The extra purple diamonds show S gates that are applied after every CZ (and eventually recompiled together with the nondisplayed layer of single-qubit gates) that map the Pauli operators P_{YY} and P_{XX} into each other. (c) In this case, CB protocols can only determine with high accuracy the product of eigenvalues $f_{XI}f_{ZX}$, which cannot be individually measured with high accuracy.

that the Pauli eigenvalues corresponding to the elements of each set therefore cannot be determined individually with high accuracy. However, it is important to stress that, out of the eight DOF associated with these eight eigenvalues, the CB protocols still allow us to determine six independent high-accuracy constraints on the products of eigenvalue pairs, thus leaving us with only two unlearnable DOF. Indeed, focusing on set $\mathcal{S}_A$, one can determine the product $f_{IX}f_{ZX}$, as shown in Fig. 8(c), as well as $f_{IY}f_{ZY}$ and $f_{IX}f_{ZY}$ [the fourth product is not independent since $f_{IY}f_{ZX} = (f_{IY}f_{ZY})(f_{IX}f_{ZX})/(f_{IX}f_{ZY})$]. The same applies to the other set $\mathcal{S}_B$. All in all, this example shows that a single CZ gate is associated with two unlearnable DOF, which cannot be determined using CB protocols and which can be (nonunambiguously) identified with eigenvalues f_{IX} and f_{XI} . These results mirror the findings discussed in Ref. [22] for a single cnot gate.

APPENDIX C: EQUIVALENCE BETWEEN PRODUCTS OF EIGENVALUES

The objective of this section is to establish the equivalence between products of eigenvalues directly measurable via MLCB—such as o_3, o_4, o'_4, o_5, c_4 introduced in the main text—and specific ratios of unlearnable eigenvalues, $\mu_q^{(BG)}$. As established at the end of Sec. IV A, two functions of eigenvalues are deemed equivalent if one can be fully derived from the other using only high-accuracy data, i.e., (products of) learnable eigenvalues.

To achieve this, we first introduce an equivalence test to formally prove the equivalence between functions of eigenvalues within a single layer. We then extend this test to multilayer scenarios and demonstrate its application using a concrete example of three-qubit chains. This example illustrates each step required to prove equivalence and establish relationships such as those expressed in Eq. (15). Building on this example, we generalize the methodology to more complex cases and summarize the results.

1. Equivalence test

We begin by focusing on a single layer C and linear combinations of logarithms of eigenvalues. These combinations can be written as $\zeta_Y = \sum_{\alpha \in \mathcal{X}_Y} \gamma_\alpha \log f_\alpha^C$, where γ_α^Y are numerical coefficients and $\mathcal{X}_Y$ is a set of Pauli strings.

Within the SPL model, the vector of model parameters $\vec{\lambda}$ determines all functions of eigenvalues via Eq. (6). In particular, the vector $\vec{\Xi}$, consisting of all learnable (products of) eigenvalues that can be determined, for instance, by following the strategy outlined in Sec. III A, is determined by

$$\log(\vec{\Xi}) = \vec{M}\vec{\lambda}. \quad (C1)$$

Analogously, ζ_Y can be determined from the model parameters via

$$\zeta_Y = (\vec{m}_Y)^T \vec{\lambda}, \quad (C2)$$

where the κ th element of the single-row matrix $(\vec{m}_Y)^T$ reads

$$(\vec{m}_Y)_\kappa = \sum_{\alpha \in \mathcal{X}_Y} \gamma_\alpha \langle \alpha, \kappa \rangle_{sp}. \quad (C3)$$

To test whether two functions of eigenvalues, ζ_A and ζ_B , are equivalent, we compute $\vec{M}$, $\vec{m}_A$, and $\vec{m}_B$. Let r be the rank of $\vec{M}$, r_Y the rank of matrix $\vec{M}$ concatenated with the extra row $(\vec{m}_Y)^T$, and r_{AB} the rank of the matrix $\vec{M}$ concatenated with $(\vec{m}_A)^T$ and $(\vec{m}_B)^T$. The equivalence condition is then

$$\zeta_A \text{ and } \zeta_B \text{ are equivalent} \Leftrightarrow r_A = r_B = r_{AB}. \quad (C4)$$

Indeed, $r_A = r_{AB}$ shows that $(\vec{m}_B)^T$ is a linear combination of the rows of $\vec{M}$ and of $(\vec{m}_A)^T$, meaning that ζ_B can be expressed in terms of ζ_A and some learnable quantities in $\vec{\Xi}$. The same logic applies symmetrically for A and B .

Notably, if $r_Y = r$, then ζ_Y is fully learnable and provides no information for fitting the model parameters.

2. Three-qubit open chain

We now consider the three-qubit open chain described in Sec. IV A. Our goals are (1) to prove the equivalence between $o_3 = f_{XIX}^B f_{XZX}^G$ and $\mu_1^{(BG)} = f_{XII}^B / f_{IIX}^G$ (2) to prove Eq. (15), which directly links these quantities.

For concreteness, we assume a QPU with square topology, ensuring that qubits 0 and 2 are not neighbors. In this case, the SPL model describing the noise in the chain features a total of $|\mathcal{K}| = 27$ parameters, associated with the Pauli strings κ listed on the x axes of the plots in Fig. 9.

We first analyze the blue layer, focusing on the function of (now trivial) eigenvalues $F_1^B = f_{XII}^B$ and $F_2^B = f_{XIX}^B$. Computing the corresponding matrix rows

$$(\vec{m}_{XII}^T)_\kappa = \langle XII, \kappa \rangle_{sp}, \quad (C5)$$

$$(\vec{m}_{XIX}^T)_\kappa = \langle XIX, \kappa \rangle_{sp}, \quad (C6)$$

we verify the equivalence condition from Eq. (C4). This proves that $\log(F_1^B)$ can be expressed as a linear combination of $\log(F_2^B)$ and other functions of learnable eigenvalues. To explicitly identify the latter, we implement a simple search algorithm, which progressively attempts to remove (randomly chosen) rows of matrix $\vec{M}$ while still preserving the condition on the ranks in Eq. (C4). We find that, in this case, the single

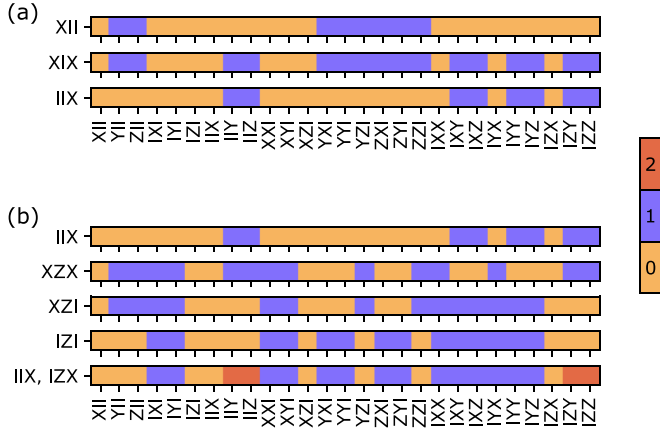


FIG. 9. Visualization of single-row matrices $(\tilde{m}_\alpha)^T$ associated with the blue layer (panel a) and the green layer (panel b) for the three-qubit open chain depicted in Fig. 3(c). On the horizontal axes, all the 27 ordered Pauli strings $\kappa \in \mathcal{K}$ are listed. The first nine are weight-one Pauli strings, the second nine are weight-two Pauli strings with support on qubits 0 and 1, and the last nine are weight-two Pauli strings with support on qubits 1 and 2. As for the Pauli strings α , they are specified to the left of each row. Note that the last row actually depicts $(\tilde{m}_j^T) = (\tilde{m}_{IIX}^T) + (\tilde{m}_{IZX}^T)$.

row of matrix $\tilde{M}$ associated with the learnable eigenvalue f_{IIX}^B , i.e.,

$$(\tilde{m}_{IIX}^T)_\kappa = \langle IIX, \kappa \rangle_{sp}, \quad (C7)$$

is sufficient to relate $(\tilde{m}_{XII}^T)$ and $(\tilde{m}_{XIX}^T)$ via $(\tilde{m}_{XII}^T) = (\tilde{m}_{XIX}^T) - (\tilde{m}_{IIX}^T)$. This is visualized in Fig. 9(a). In view of the subsequent analysis, it is convenient to summarize this result as

$$\log(F_1^L) = \epsilon^L \log(F_2^L) + (\tilde{\Sigma}^L)^T \log(\tilde{F}^L), \quad (C8)$$

with $L = B$, $\epsilon^B = 1$, $\tilde{\Sigma}^B = (-1)$, and $\tilde{F}^B = (f_{IIX}^B)$. This proves

$$f_{XII}^B = \frac{f_{XIX}^B}{f_{IIX}^B}. \quad (C9)$$

Similarly, for the green layer we focus on functions $F_1^G = f_{IIX}^G$ and $F_2^G = f_{XZX}^G$, proven to be equivalent by Eq. (C4). Moreover, we show that the corresponding matrix rows $(\tilde{m}_{IIX}^T)$ and $(\tilde{m}_{XZX}^T)$ can be related to each other by using three rows of matrix $\tilde{M}$. Specifically, we obtained

$$(\tilde{m}_{IIX}^T) = -(\tilde{m}_{XZX}^T) + (\tilde{m}_{XZI}^T) - (\tilde{m}_{IZI}^T) + (\tilde{m}_j^T), \quad (C10)$$

where the last row $(\tilde{m}_j^T) = (\tilde{m}_{IIX}^T) + (\tilde{m}_{IZX}^T)$ is associated with the learnable product of eigenvalues in the orbit $\mathcal{R}_{IIX}^G = \{P_{IIX}, P_{IZX}\}$. Such a relation is visualized in Fig. 9(b). This result can be summarized in the form of Eq. (C8), with $L = G$, $\epsilon^G = -1$, $\tilde{\Sigma}^G = (1, -1, 1)$, and $\tilde{F}^G = (f_{XZI}^G, f_{IZI}^G, f_{IIX}^G f_{IZX}^G)$. This proves

$$f_{IIX}^G = \frac{f_{XZI}^G (f_{IZX}^G f_{IIX}^G)}{f_{XZX}^G f_{IZI}^G}. \quad (C11)$$

By combining these results, valid for each layer individually, we obtain Eq. (15), thus showing the equivalence

between $o_3 = F_2^B F_2^G$ and $\mu_1^{(BG)} = F_1^B / F_1^G$. It is important to stress that the equivalence class does not consist only of these two functions of eigenvalues. It is indeed possible to find other equivalent functions of eigenvalues, all associated with the same DOF, whose learnability is unlocked by MLCB. For example, one has

$$f_{XII}^B f_{IZX}^G = \mu_1^{(BG)} (f_{IIX}^G f_{IZX}^G), \quad (C12)$$

where $f_{IIX}^G f_{IZX}^G$ is learnable with conventional CB.

3. Other chains

Longer chains introduce additional challenges due to the connectivity of the underlying QPU. Non-neighboring qubits in the chain may still interact via the hardware's topology, increasing the number of SPL parameters and altering the relationships between functions of eigenvalues. Relevant examples are shown in the first column of Tables II, III, IV, and V.

In the following, we analyze all relevant chain configurations, considering a QPU with a square topology. For each case, we apply the methodology outlined for the three-qubit chain. Specifically, focusing on one layer at a time, we use the equivalence test to demonstrate the equivalence between two functions of eigenvalues, denoted as F_1^L and F_2^L , respectively. This approach enables us to prove the equivalence—and derive a direct connection—between the eigenvalue product measurable via MLCB, $F_2^B F_2^G$, and the ratios of unlearnable eigenvalues, F_1^B / F_1^G . The results for all chain configurations are summarized in Tables II, III, IV, and V.

It is important to note that the reported $\tilde{\Sigma}^L$ and $\tilde{F}^L$ are not unique, as alternative choices for these quantities can still satisfy Eq. (C8). An interesting open question concerns the optimization of the choice of $\tilde{\Sigma}^L$ and $\tilde{F}^L$, with the goal of minimizing the number of components in each vector for a given chain. While this optimization does not affect the learnability of eigenvalues and is thus beyond the scope of this paper, it can be beneficial when implementing the characterization procedure in practice. For instance, in the four-qubit closed chain analyzed in Table V, connecting $c_4 = f_{IXXX}^B f_{ZXXY}^G f_{IYYX}^B f_{ZYXY}^G$ to the simple eigenvalue ratio $\mu_q^{(BG)}$ requires considering a total of 28 learnable (products of) eigenvalues. Reducing this number could streamline the analyses and make them more robust with respect to the spread of statistical uncertainties.

APPENDIX D: EXPERIMENTAL DETAILS

In this Appendix, we provide further details on the experimental execution and data analysis of the MLCB protocol discussed in Sec. V.

For each of the six pairs of layers (L_1, L_2) depicted in Fig. 5(a), we perform a complete MLCB experiment. Each experiment consists of running circuits for five different sequence depths $d \in 2, 4, 8, 16, 32$. For each depth, we use a total of 4×10^4 measurement shots, which are distributed evenly across 20 random Pauli twirling instances. This results in 2000 shots per unique circuit configuration. To gather more statistics and develop a feeling of the time stability of the results, we repeat this entire procedure three times back-to-back for every layer pair.

TABLE II. Four-qubit open chains (blue-green-blue).

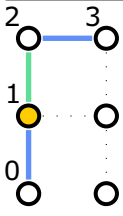
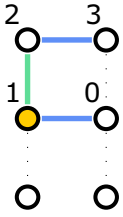
	F_1^B	F_2^B	ϵ^B	$\bar{\Sigma}^B$	$\bar{F}^B$
	f_{XIII}^B	$f_{XIXI}^B f_{XIXZ}^B$	+0.50	(-0.50)	$(f_{IIXZ}^B f_{IIXI}^B)$
	F_1^G	F_2^G	ϵ^G	$\bar{\Sigma}^G$	$\bar{F}^G$
	f_{IIXI}^G	$f_{XZXZ}^G f_{XZXI}^G$	-0.50	$\begin{pmatrix} +0.50 \\ -0.50 \\ +0.25 \\ +0.75 \end{pmatrix}$	$\begin{pmatrix} f_{XZII}^G f_{XZII}^G \\ f_{IZII}^G f_{IZII}^G \\ f_{IIXZ}^G f_{IIXZ}^G \\ f_{IZXI}^G f_{IIXI}^G \end{pmatrix}$
	F_1^B	F_2^B	ϵ^B	$\bar{\Sigma}^B$	$\bar{F}^B$
	f_{XIII}^B	$f_{XIXI}^B f_{XIXZ}^B$	+0.50	$\begin{pmatrix} +0.25 \\ -0.50 \\ +0.25 \\ -0.25 \end{pmatrix}$	$\begin{pmatrix} f_{IIIZ}^B f_{IIIZ}^B \\ f_{IIXZ}^B f_{IIXI}^B \\ f_{XZII}^B f_{XIII}^B \\ f_{XIIZ}^B f_{XZIZ}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\bar{\Sigma}^G$	$\bar{F}^G$
	f_{IIXI}^G	$f_{XZXZ}^G f_{XZXI}^G$	-0.50	$\begin{pmatrix} -0.50 \\ +0.25 \\ +0.75 \\ -0.25 \\ +0.50 \\ -0.25 \\ +0.25 \end{pmatrix}$	$\begin{pmatrix} f_{IZII}^G f_{IZII}^G \\ f_{IIXZ}^G f_{IIXZ}^G \\ f_{IIXI}^G f_{IZXI}^G \\ f_{IIIZ}^G f_{IIIZ}^G \\ f_{XZII}^G f_{XZII}^G \\ f_{XIII}^G f_{XIII}^G \\ f_{XIIZ}^G f_{XIIZ}^G \end{pmatrix}$

TABLE III. Four-qubit open chains (green-blue-green).

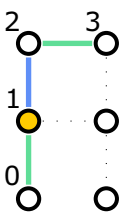
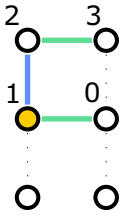
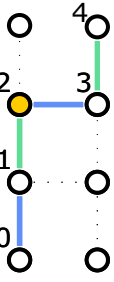
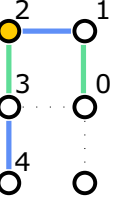
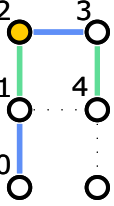
	F_1^B	F_2^B	ϵ^B	$\bar{\Sigma}^B$	$\bar{F}^B$
	f_{IXII}^B	$f_{IXIX}^B f_{ZXIX}^B$	+0.50	$\begin{pmatrix} -0.25 \\ +0.25 \\ -0.50 \end{pmatrix}$	$\begin{pmatrix} f_{ZXII}^B f_{ZXZI}^B \\ f_{IXZI}^B f_{IXII}^B \\ f_{IIXX}^B f_{IIXX}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\bar{\Sigma}^G$	$\bar{F}^G$
	f_{IIXI}^G	$f_{XZXZ}^G f_{XZXI}^G$	-0.50	$\begin{pmatrix} -0.50 \\ +0.50 \\ +1.00 \end{pmatrix}$	$\begin{pmatrix} f_{IIZI}^G f_{IIZI}^G \\ f_{IXZI}^G f_{ZXZI}^G \\ f_{IIZX}^G f_{IIXX}^G \end{pmatrix}$
	F_1^B	F_2^B	ϵ^B	$\bar{\Sigma}^B$	$\bar{F}^B$
	f_{IXII}^B	$f_{IXIX}^B f_{ZXIX}^B$	+0.50	$\begin{pmatrix} +0.25 \\ -0.25 \\ -0.25 \\ +0.25 \\ -0.25 \end{pmatrix}$	$\begin{pmatrix} f_{ZIII}^B f_{ZIII}^B \\ f_{IIXX}^B f_{IIXX}^B \\ f_{ZXII}^B f_{ZXZI}^B \\ f_{IXZI}^B f_{IXII}^B \\ f_{ZIIIX}^B f_{ZIIIX}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\bar{\Sigma}^G$	$\bar{F}^G$
	f_{IIXI}^G	$f_{XZXZ}^G f_{XZXI}^G$	-0.50	$\begin{pmatrix} -0.25 \\ +0.25 \\ +0.50 \\ -0.50 \\ +0.75 \end{pmatrix}$	$\begin{pmatrix} f_{ZIII}^G f_{ZIII}^G \\ f_{ZIIIX}^G f_{ZIZX}^G \\ f_{IXZI}^G f_{ZXZI}^G \\ f_{IIZI}^G f_{IIZI}^G \\ f_{IIZX}^G f_{IIXX}^G \end{pmatrix}$

TABLE IV. Five-qubit open chains.

	F_1^B	F_2^B	ϵ^B	$\tilde{\Sigma}^B$	$\tilde{F}^B$
	f_{IXIII}^B	$f_{IXIXI}^B f_{ZXIXZ}^B$	+0.50	$\begin{pmatrix} -0.25 \\ +0.25 \\ -0.50 \end{pmatrix}$	$\begin{pmatrix} f_{ZXIII}^B f_{ZXZII}^B \\ f_{IXIXI}^B f_{IXZII}^B \\ f_{IIIIXI}^B f_{IIIXZ}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\tilde{\Sigma}^G$	$\tilde{F}^G$
	f_{IIIXI}^G	$f_{IXZXZ}^G f_{ZXZXI}^G$	-0.50	$\begin{pmatrix} -0.50 \\ +0.75 \\ +0.25 \\ +0.50 \end{pmatrix}$	$\begin{pmatrix} f_{IIIZII}^G f_{IIIZII}^G \\ f_{IIIZXI}^G f_{IIIXI}^G \\ f_{IIIXZ}^G f_{IIIZXZ}^G \\ f_{IXZII}^G f_{ZXZII}^G \end{pmatrix}$
	F_1^B	F_2^B	ϵ^B	$\tilde{\Sigma}^B$	$\tilde{F}^B$
	f_{IXIII}^B	$f_{IXIXI}^B f_{ZXIXZ}^B$	+0.50	$\begin{pmatrix} +0.25 \\ -0.25 \\ -0.25 \\ +0.25 \\ -0.25 \end{pmatrix}$	$\begin{pmatrix} f_{IXZII}^B f_{IXIII}^B \\ f_{ZIIIXI}^B f_{ZIIIXZ}^B \\ f_{ZXIII}^B f_{ZXZII}^B \\ f_{ZIIII}^B f_{ZIIII}^B \\ f_{IIIXZ}^B f_{IIIXI}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\tilde{\Sigma}^G$	$\tilde{F}^G$
	f_{IIIXI}^G	$f_{IXZXZ}^G f_{ZXZXI}^G$	-0.50	$\begin{pmatrix} +0.25 \\ +0.25 \\ -0.50 \\ +0.50 \\ -0.25 \\ +0.50 \end{pmatrix}$	$\begin{pmatrix} f_{IIIXZ}^G f_{IIIZXZ}^G \\ f_{ZIIIXI}^G f_{ZIZXI}^G \\ f_{IIIZII}^G f_{IIIZII}^G \\ f_{IXZII}^G f_{ZXZII}^G \\ f_{ZIIII}^G f_{ZIIII}^G \\ f_{IIIXZ}^G f_{IIIXI}^G \end{pmatrix}$
	F_1^B	F_2^B	ϵ^B	$\tilde{\Sigma}^B$	$\tilde{F}^B$
	f_{IXIII}^B	$f_{IXIXI}^B f_{ZXIXZ}^B$	+0.50	$\begin{pmatrix} +0.50 \\ -0.50 \\ -0.25 \\ -0.25 \\ +0.25 \end{pmatrix}$	$\begin{pmatrix} f_{IXZII}^B f_{IXIII}^B \\ f_{IIIXI}^B f_{IIIXZ}^B \\ f_{IXIIZ}^B f_{IXZIZ}^B \\ f_{ZXIII}^B f_{ZXZII}^B \\ f_{IIIZ}^B f_{IIIZ}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\tilde{\Sigma}^G$	$\tilde{F}^G$
	f_{IIIXI}^G	$f_{IXZXZ}^G f_{ZXZXI}^G$	-0.50	$\begin{pmatrix} +0.25 \\ +0.50 \\ -0.25 \\ +0.25 \\ +0.75 \\ -0.25 \\ -0.50 \end{pmatrix}$	$\begin{pmatrix} f_{IXIIZ}^G f_{ZXIIZ}^G \\ f_{IXZII}^G f_{ZXZII}^G \\ f_{ZXIII}^G f_{IXIII}^G \\ f_{IIIXZ}^G f_{IIIZXZ}^G \\ f_{IIIXI}^G f_{IIIZXI}^G \\ f_{IIIZ}^G f_{IIIZ}^G \\ f_{IIIZII}^G f_{IIIZII}^G \end{pmatrix}$

The statistical uncertainty of our measurements is estimated using a bootstrapping procedure. For each of the three experimental runs, we generate 100 bootstrap samples by resampling with replacement from both the measurement shots and the 20 random twirling instances. The final reported values and confidence intervals for $\Delta_q^{L_1 L_2}$ are derived from the combined distribution of these bootstrapped results, averaged across the three consecutive runs. In Fig. 10, we present the complete set of measured $\Delta_q^{L_1 L_2}$ values for all six layer pairs, averaged over the three consecutive experimental runs.

To assess the normality of the bootstrapped distributions, we compute the skewness s_k and excess kurtosis k_u for each difference $\Delta_q^{L_1 L_2}$. With the exception of Δ_5^{OR} , which exhibits large skewness $s_k = -1$ and excess kurtosis $k_u = 2.5$ (as also evident from the asymmetric whiskers in Fig. 10), all the other differences features small values within the ranges $-0.18 \leq s_k \leq 0.25$ and $-0.14 \leq k_u \leq 0.32$. This confirms the approximate normality of the bootstrapped distributions, thereby validating the use of the associated ratios $\Omega_q^{L_1 L_2}$, between mean and standard deviations, to assess their statistical compatibility with zero.

TABLE V. Four-qubit closed chain.

	F_1^B	F_2^B	ϵ^B	$\bar{\Sigma}^B$	$\bar{F}^B$
	f_{IXII}^B	$f_{IXXX}^B f_{IYYX}^B$	+0.50	$\begin{pmatrix} +0.25 \\ -0.50 \\ -0.25 \\ -0.25 \\ -0.25 \\ -0.25 \\ -0.25 \\ +0.50 \\ +0.25 \\ +0.25 \\ +0.50 \\ -0.25 \end{pmatrix}$	$\begin{pmatrix} f_{IIXY}^B f_{IIXY}^B \\ f_{IIXY}^B f_{IIYX}^B \\ f_{IXIX}^B f_{ZXZX}^B \\ f_{ZIII}^B f_{ZIII}^B \\ f_{IIXX}^B f_{IIXX}^B \\ f_{IYIX}^B f_{ZYZX}^B \\ f_{IIZI}^B f_{IIZI}^B \\ f_{IIZX}^B f_{IIZX}^B \\ f_{ZIZI}^B f_{ZIZI}^B \\ f_{IXII}^B f_{ZXII}^B \\ f_{IXII}^B f_{ZYII}^B \\ f_{ZYII}^B f_{ZYII}^B \end{pmatrix}$
	F_1^G	F_2^G	ϵ^G	$\bar{\Sigma}^G$	$\bar{F}^G$
	$f_{II XI}^G$	$f_{ZXY Y}^G f_{ZYXY}^G$	-0.50	$\begin{pmatrix} -0.25 \\ +0.25 \\ +0.25 \\ -0.25 \\ +0.25 \\ +0.25 \\ +0.50 \\ -0.25 \\ -0.25 \\ +0.25 \\ +0.50 \\ -0.25 \\ +0.25 \\ -0.25 \\ -0.25 \\ +0.25 \end{pmatrix}$	$\begin{pmatrix} f_{ZZII}^G f_{ZZII}^G \\ f_{IXIY}^G f_{IXIY}^G \\ f_{IIXY}^G f_{ZZXY}^G \\ f_{IYIY}^G f_{ZIZI}^G \\ f_{ZXII}^G f_{ZXIZ}^G \\ f_{ZYII}^G f_{ZYIZ}^G \\ f_{IXIX}^G f_{IYIY}^G \\ f_{IXIZ}^G f_{IXII}^G \\ f_{ZIII}^G f_{ZIII}^G \\ f_{IIZI}^G f_{IIZI}^G \\ f_{IIXI}^G f_{ZIZI}^G \\ f_{IXIX}^G f_{IXIX}^G \\ f_{ZIXI}^G f_{IIXI}^G \\ f_{IIYI}^G f_{IZIY}^G \\ f_{IYII}^G f_{IYIZ}^G \\ f_{IYIY}^G f_{ZZYI}^G \end{pmatrix}$

As detailed in Fig. 11(a), the products of Pauli eigenvalues measured by MLCB deviate from 1 by around 0.1. In contrast, the absolute differences $|\Delta_q^{L_1 L_2}|$ remain below 0.006 [see Figs. 10 and 11(b)]. While these deviations are statistically significant, as established above, their magnitude indicates that violations of the symmetry assumption between unlearnable quantities contribute only a few percent to the overall value of the Pauli eigenvalues.

Let us now illustrate with a detailed example what is accessible to MLCB and standard CB protocols in terms of products of Pauli eigenvalues. To this end, we focus on the layers' pair BR and on qubit $q = 11$, i.e., to the case analyzed in detail in Fig. 5(c). Here, we have

$$\mathcal{F}_{\alpha(11)}^{BR} = f_{IXIX}^B f_{IXZX}^R f_{ZXIX}^B f_{ZXZX}^R, \quad (D1)$$

where the Pauli strings refer to qubits 9, 10, 11, and 12 [see Fig. 5(a)]. As $\beta(11) = IXZX$, we also have

$$\mathcal{F}_{\beta(11)}^{BR} = f_{IXZX}^B f_{ZXIX}^R f_{ZXZX}^B f_{IXIX}^R. \quad (D2)$$

Those two products can be measured with high accuracy with MLCB but cannot be individually accessed via conventional

CB. The latter could only be used to measure, by combining the analyses of the two layers, the product $f = \mathcal{F}_{\alpha(11)}^{BR} \mathcal{F}_{\beta(11)}^{BR}$ that reads

$$f = (f_{IXIX}^B f_{IXZX}^B)(f_{ZXIX}^B f_{ZXZX}^B)(f_{IXIX}^R f_{ZXZX}^R)(f_{IXZX}^R f_{ZXIX}^R) \quad (D3)$$

and all the products within round brackets are learnable with standard CB.

APPENDIX E: FITTING THE NOISE PARAMETERS

Here, we detail how we impose the high-accuracy constraint obtained with MLCB on the low-accuracy eigenvalues. Let us focus on a qubit q , belonging to the support of l layers $L_1, \dots, L_l$. When relying exclusively on conventional CB protocols, qubit q is associated with l unlearnable DOF, which can be identified with the l eigenvalues in $\{f_{\beta_q^j}^{L_j}\}$, where the weight-one Pauli strings β_q^j feature a single X operator on the qubit connected to q via the L_j layer. Low-accuracy methods, like the symmetry assumption or unit-depth

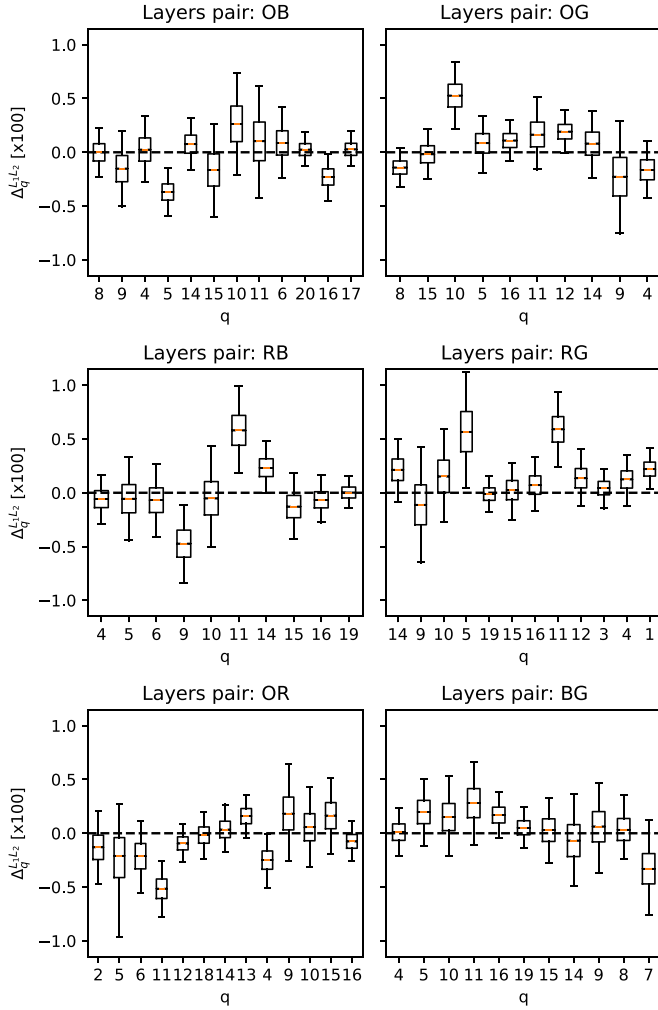


FIG. 10. Summary of all measured asymmetry values $\Delta_q^{L_1 L_2}$. Each panel corresponds to a different pair of layers, as indicated by the title. The box plots show the distribution of Δ_q for each hot qubit q in the chain, averaged over three consecutive MLCB experiments. The whiskers indicate the 95% confidence interval derived from bootstrapping.

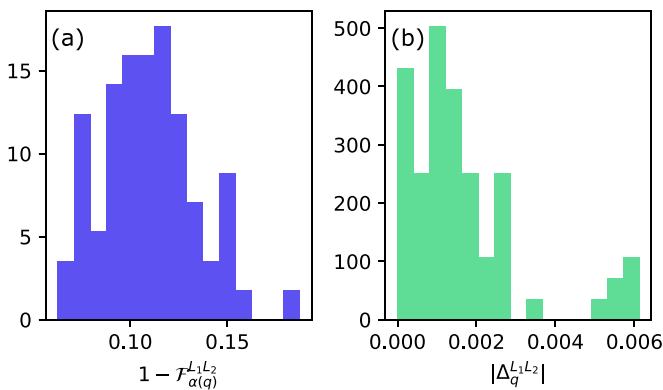


FIG. 11. Distribution of all the ‘infidelities’ $1 - \mathcal{F}_{\alpha(q)}^{L_1 L_2}$ (panel a) and of all the absolute differences $|\Delta_q^{L_1 L_2}|$ (panel b) measured by the MLCB protocol, for all q and all layer pairs.

CB-like experiments, provide us with estimates for each one of these eigenvalues, which we indicate with an extra tilde $\tilde{f}_{\beta_q}^{L_j}$ and represent the numerical values used for the conventional model fitting.

MLCB allows us to determine $l - 1$ additional high-accuracy constraints, which can be identified, for example, with eigenvalue ratios

$$\mu_q^{L_j, L_{j+1}} = \frac{\tilde{f}_{\beta_q}^{L_j}}{\tilde{f}_{\beta_q}^{L_{j+1}}} \quad (\text{E1})$$

for $j = 1, \dots, l - 1$. This reduces the number of unlearnable DOF to one, which we can identify with a specific eigenvalue, say $\tilde{f}_q^{L_1}$. We estimate its numerical value u by minimizing the sum S of the squared residues

$$S = (\tilde{f}_q^{L_j} - u)^2 + \sum_{j=1}^{l-1} \left(\tilde{f}_{\beta_q}^{L_j} - u \prod_{i=1}^j \mu_q^{L_i, L_{i+1}} \right)^2, \quad (\text{E2})$$

computed by comparing the low-accuracy eigenvalue estimations with the effect of the high-accuracy constraints from MLCB. This procedure can be readily adjusted to tackle scenarios where different pairs of layers have been studied with MLCB.

APPENDIX F: RANDOM NOISE MODELS

The random noise models are generated to simulate the behavior of realistic devices. The procedure involves assigning values to all model parameters for a given layer L distinguishing them based on the following criteria:

- (1) Weight (w) of the Pauli operator P_κ , either $w = 1$ or $w = 2$;
- (2) Status (s) of the involved qubits, which can be either active ($s = a$) if all qubits in the support of P_κ belong to the support of L , or inactive ($s = i$) otherwise.

Inactive parameters are drawn from a Gaussian distribution with mean m_w^i and standard deviation σ_w^i . The active ones are sampled from a Gaussian with mean m_w^a and standard deviation σ_w^a , where, for every gate in the layer, m_w^a itself is drawn from a Gaussian with mean $\bar{m}_w^a$ and standard deviation $\bar{\sigma}_w^a$. This additional sampling step introduces variability between gates within a single layer, resulting in better and worse-performing ones.

Finally, note that any negative model parameters resulting from the sampling process are set to zero. The parameter values used in this procedure read

$$\begin{aligned} m_1^i &= 2 \times 10^{-4} & \sigma_1^i &= 8 \times 10^{-4} \\ \bar{m}_1^a &= 1 \times 10^{-3} & \bar{\sigma}_1^a &= 7.5 \times 10^{-4} \\ & & \sigma_1^a &= 1 \times 10^{-3} \end{aligned} \quad (\text{F1})$$

for $w = 1$ and

$$\begin{aligned} m_2^i &= 1.5 \times 10^{-4} & \sigma_2^i &= 1 \times 10^{-3} \\ \bar{m}_2^a &= 2 \times 10^{-3} & \bar{\sigma}_2^a &= 1.5 \times 10^{-3} \\ & & \sigma_2^a &= 2 \times 10^{-3} \end{aligned} \quad (\text{F2})$$

for $w = 2$.

- [1] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nat. Commun.* **8**, 14485 (2017).
- [2] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
- [3] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, *Nat. Phys.* **14**, 595 (2018).
- [4] J. Chen, D. Ding, C. Huang, and L. Kong, Linear cross-entropy benchmarking with Clifford circuits, *Phys. Rev. A* **108**, 052613 (2023).
- [5] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Scalable randomized benchmarking of quantum computers using mirror circuits, *Phys. Rev. Lett.* **129**, 150502 (2022).
- [6] J. Hines, M. Lu, R. K. Naik, A. Hashim, J.-L. Ville, B. Mitchell, J. M. Kriekebaum, D. I. Santiago, S. Seritan, E. Nielsen, R. Blume-Kohout, K. Young, I. Siddiqi, B. Whaley, and T. Proctor, Demonstrating scalable randomized benchmarking of universal gate sets, *Phys. Rev. X* **13**, 041030 (2023).
- [7] J. Hines, D. Hothem, R. Blume-Kohout, B. Whaley, and T. Proctor, Fully scalable randomized benchmarking without motion reversal, *PRX Quantum* **5**, 030334 (2024).
- [8] 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).
- [9] S. T. Flammia, Averaged circuit eigenvalue sampling, in *Proceedings of the 17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, edited by F. L. Gall and T. Morimae, Leibniz International Proceedings in Informatics (LIPIcs), Dagstuhl, Germany (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2022), Vol. 232, pp. 4:1–4:10, <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.TQC.2022.4>.
- [10] 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*.
- [11] S. Mangini, M. Cattaneo, D. Cavalcanti, S. Filippov, M. A. C. Rossi, and G. García-Pérez, Tensor network noise characterization for near-term quantum computers, *Phys. Rev. Res.* **6**, 033217 (2024).
- [12] 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).
- [13] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [14] S. Filippov, M. Leahy, M. A. C. Rossi, and G. García-Pérez, Scalable tensor-network error mitigation for near-term quantum computing, *arXiv:2307.11740*.
- [15] 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).
- [16] E. T. Hockings, A. C. Doherty, and R. Harper, Improving error suppression with noise-aware decoding, *arXiv:2502.21044*.
- [17] K. Tiurev, P.-J. H. S. Derks, J. Roffe, J. Eisert, and J.-M. Reiner, Correcting non-independent and non-identically distributed errors with surface codes, *Quantum* **7**, 1123 (2023).
- [18] 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*.
- [19] O. Kern, G. Alber, and D. L. Shepelyansky, Quantum error correction of coherent errors by randomization, *Eur. Phys. J. D* **32**, 153 (2005).
- [20] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [21] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kriekebaum, M. Davis, E. Smith, C. Iancu, K. P. O’Brien, *et al.*, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, *Phys. Rev. X* **11**, 041039 (2021).
- [22] S. Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, and L. Jiang, The learnability of Pauli noise, *Nat. Commun.* **14**, 52 (2023).
- [23] 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).
- [24] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, *Phys. Rev. A* **87**, 062119 (2013).
- [25] 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).
- [26] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, *ACM Trans. Quantum Comput.* **1**, 1 (2020).
- [27] L. Abdurakhimov, J. Adam, H. Ahmad, O. Ahoenen, M. Algaba, G. Alonso, V. Bergholm, R. Beriwal, M. Beuerle, C. Bockstiegel, *et al.*, Technology and performance benchmarks of IQM’s 20-qubit quantum computer, *arXiv:2408.12433*.
- [28] 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).
- [29] 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).
- [30] A. W. R. Smith, K. E. Khosla, C. N. Self, and M. S. Kim, Qubit readout error mitigation with bit-flip averaging, *Sci. Adv.* **7**, eabi8009 (2021).
- [31] A. Hashim, A. Carignan-Dugas, L. Chen, C. Juenger, N. Fruitwala, Y. Xu, G. Huang, J. J. Wallman, and I. Siddiqi, Quasi-probabilistic readout correction of mid-circuit measurements for adaptive feedback via measurement randomized compiling, *PRX Quantum* **6**, 010307 (2025).
- [32] S. J. Beale and J. J. Wallman, Randomized compiling for subsystem measurements, *arXiv:2304.06599*.
- [33] R. Harper, I. Hincks, C. Ferrie, S. T. Flammia, and J. J. Wallman, Statistical analysis of randomized benchmarking, *Phys. Rev. A* **99**, 052350 (2019).

- [34] B. McDonough, A. Mari, N. Shammah, N. T. Stemen, M. Wahl, W. J. Zeng, and P. P. Orth, Automated quantum error mitigation based on probabilistic error reduction, in *Proceedings of the IEEE/ACM Third International Workshop on Quantum Computing Software (QCS)* (IEEE/ACM, 2022), pp. 83–93.
- [35] L. E. Fischer, M. Leahy, A. Eddins, N. Keenan, D. Ferracin, M. A. C. Rossi, Y. Kim, A. He, F. Pietracaprina, B. Sokolov, *et al.*, Dynamical simulations of many-body quantum chaos on a quantum computer, [arXiv:2411.00765](#).
- [36] S. N. Filippov, S. Maniscalco, and G. García-Pérez, Scalability of quantum error mitigation techniques: From utility to advantage, [arXiv:2403.13542](#).
- [37] Implicitly, this means that we define an ordering of the set $\mathcal{K}$, which maps a Pauli string α to an integer $i \in \mathbb{N}$ so that the i th element of $\tilde{\lambda}$ (denoted by λ_i) is given by λ_α^C .
- [38] A. Mari, N. Shammah, and W. J. Zeng, Extending quantum probabilistic error cancellation by noise scaling, *Phys. Rev. A* **104**, 052607 (2021).
- [39] This can be done by randomly sampling circuits containing Pauli errors P_k according to the (quasi)probability distribution defined by coefficients $w_k^{(\beta)}$ [8].
- [40] Google Quantum AI., Suppressing quantum errors by scaling a surface code logical qubit, *Nature (London)* **614**, 676 (2023).
- [41] 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](#).
- [42] Indeed, it is required that each one of the two elements of β in the support of α is either the identity or equal to the corresponding element of α . This can be always ensured by composing the layer C with a second Clifford layer, consisting only of single-qubit gates [8].
- [43] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [44] T. E. O’Brien, G. Anselmetti, F. Gkritis, V. E. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, *et al.*, Purification-based quantum error mitigation of pair-correlated electron simulations, *Nat. Phys.* **19**, 1787 (2023).
- [45] This result is not limited to square lattices but applies more generally to topologies where qubits can be ordered such that no qubit is directly connected to more than two of its predecessors.
- [46] E. van den Berg and P. Wocjan, Techniques for learning sparse Pauli-Lindblad noise models, *Quantum* **8**, 1556 (2024).
- [47] S. Chen, C. Oh, S. Zhou, H.-Y. Huang, and L. Jiang, Tight bounds on Pauli channel learning without entanglement, *Phys. Rev. Lett.* **132**, 180805 (2024).
- [48] N. Fruitwala, A. Hashim, A. D. Rajagopala, Y. Xu, J. Hines, R. K. Naik, I. Siddiqi, K. Klymko, G. Huang, and K. Nowrouzi, Hardware-efficient randomized compiling, [arXiv:2406.13967](#).
- [49] C. Granade, C. Ferrie, and D. G. Cory, Accelerated randomized benchmarking, *New J. Phys.* **17**, 013042 (2015).
- [50] S. Chen, S. Zhou, A. Seif, and L. Jiang, Quantum advantages for Pauli channel estimation, *Phys. Rev. A* **105**, 032435 (2022).
- [51] Y. Chen, Z. Yu, C. Zhu, and X. Wang, Efficient information recovery from Pauli noise via classical shadow, [arXiv:2305.04148](#).
- [52] S. Chen, Z. Zhang, L. Jiang, and S. T. Flammia, Efficient self-consistent learning of gate set Pauli noise, *PRX Quantum* **7**, 010305 (2026).