

Efficient Characterization of Coherent and Correlated Low-Degree Noise in Layers of Gates

Marianna Crupi^{1,2,*}, J. Ignacio Cirac^{1,2} and Flavio Baccari³

¹*Max Planck Institute of Quantum Optics, 85748 Garching, Germany*

²*Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany*

³*Dipartimento di Fisica e Astronomia “G. Galilei” & Padua Quantum Technologies Research Center, Università degli Studi di Padova, 35131 Padova, Italy*



(Received 16 July 2025; revised 23 October 2025; accepted 24 November 2025; published 30 December 2025)

We present a quantum process-tomography protocol based on a low-degree ansatz for the quantum channel, i.e., when it can be expressed as a fixed-degree polynomial in terms of Pauli operators. We demonstrate how to perform tomography of such channels with a logarithmic amount of effort relative to the size of the system, by employing random state preparation and measurements in the Pauli basis. We extend the applicability of the protocol to channels consisting of a layer of quantum gates with a polylogarithmic number of non-Clifford gates, followed by a low-degree noise channel. Rather than inverting the layer of quantum gates on the hardware—which would introduce additional errors—we instead carry out the inversion in classical postprocessing, while adding to the sample complexity a factor at most polynomial in system size. Numerical simulations support our theoretical findings and demonstrate the feasibility of our method.

DOI: [10.1103/8ng1-4c1k](https://doi.org/10.1103/8ng1-4c1k)

I. INTRODUCTION

Accurate noise characterization is a critical prerequisite for the reliable use of current quantum devices and the development of future quantum technologies. Understanding the main sources of imperfections in quantum operations is essential for designing effective error mitigation and correction techniques [1–4], as well as for assessing the capabilities of quantum hardware [5,6].

An effective noise characterization scheme should satisfy several key criteria. First, it must be scalable, in terms of both experimental implementation and the computational resources required to process the collected data. Second, the noise model should be sufficiently comprehensive to capture the complexity of realistic noise processes and ensure broad applicability across various platforms. In particular, it should account for both coherent errors and cross-talk, even with arbitrary device connectivity.

A wide range of noise characterization methods have already been developed, each navigating trade-offs between scalability and generality. Standard approaches

like quantum process tomography and gate-set tomography aim to fully reconstruct quantum channels but quickly become impractical due to their exponential resource demands [7–12]. More scalable alternatives, such as randomized benchmarking and shadow tomography, enable efficient estimation of specific noise properties but sacrifice completeness in the reconstructed information [6,13–21]. A promising intermediate strategy seeks to balance scalability and completeness by imposing assumptions on the noise model. Examples include Pauli-diagonal noise models, perturbative ansätze, or low-degree models [22–29]. While each approach suits specific use cases, they come with limitations: some require fine-tuned noise tailoring (e.g., Pauli twirling), others lack rigorous recovery guarantees, and some are experimentally impractical due to high circuit complexity.

In this work, we introduce a tomography method that maintains both scalability and completeness while overcoming the limitations of previous approaches. Our focus is on the low-degree quantum channel ansatz [29], where error events act nontrivially on at most d qubits simultaneously. This ansatz captures important physical features such as coherent errors and cross-talk, and it generalizes to devices with arbitrary connectivity. We address the scalability challenges of earlier tomography schemes [7–12] by developing a reconstruction algorithm with logarithmic sample complexity, relying only on single-qubit state preparations and measurements (SPAMs). As a result, our

*Contact author: marianna.crupi@mpq.mpg.de

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

method is both theoretically efficient and experimentally practical.

The low-degree ansatz is particularly well suited for describing weak noise channels. It can also model the noisy implementation of one or a few two-qubit gates, including potential crosstalk with idle qubits. However, when an entire layer of such gates is applied, the resulting channel ceases to be low degree, regardless of the noise strength. To address this, we model the noisy implementation as an ideal gate layer followed by a low-degree noise channel. Recovering the noise component alone requires inverting the ideal gate layer—a process that can introduce additional noise if performed on hardware. Instead, we demonstrate that in certain relevant cases, our protocol can eliminate the effect of ideal gates through classical post-processing, avoiding extra experimental overhead. This extension broadens the applicability of our protocol and makes it a versatile tool for characterizing noise in near-term quantum devices, while only adding an overhead at most polynomial in system size.

This paper is organized as follows. In Sec. II, we introduce our noise characterization framework. Section III presents the tomography protocol for low-degree channels and provides recovery guarantees. In Sec. IV, we show how to adapt the protocol to eliminate gate layers via postprocessing. Section V presents numerical results under specific noise models. Finally, Sec. VI summarizes our findings and discusses future directions.

II. PROBLEM SETTING

We consider the application of a noisy unitary on n qubits, and the goal is to completely recover the noise component. For this we assume that certain laboratory operations—specifically, state preparation and measurement—can be performed without error. In particular, we assume the ability to prepare product states of Pauli eigenstates $(\sigma_x, \sigma_y, \sigma_z)$ and to perform measurements in these bases. While this assumption may not hold in all experimental settings, it is well justified in several quantum computing platforms—for example, when high-fidelity measurements in the computational basis and accurate single-qubit Clifford gates are available.

We now recall some standard concepts from quantum process tomography, in particular those relevant for gate-error characterization, which will be useful for the development of our protocol. Let us represent the noisy implementation of a unitary by the corresponding quantum channel,

$$\mathcal{U} \mapsto \mathcal{C}_{\mathcal{U}}. \quad (1)$$

In the general case this channel can be many-body and gate-dependent, including correlated and coherent noise models.

More specifically, we would like to characterize the error component of the noisy implementation $\mathcal{C}_{\mathcal{U}}$,

$$\mathcal{E}_{\mathcal{U}} := \mathcal{C}_{\mathcal{U}} \circ \mathcal{U}^\dagger, \quad (2)$$

where we use the subscript $\mathcal{U}$ to indicate the possible gate dependence of the noise [30]. However, we will omit the subscript where there is no chance of misinterpretation.

Characterizing a generic quantum channel requires resources that scale exponentially with system size [11]. To enable an efficient representation of the noise in our implementation, we must therefore make further assumptions. We adopt a physically motivated one: the error channel is a perturbative expansion around the identity channel, corresponding to the case of no noise. More specifically, we define the Kraus operators of the error channel as in Eq. (2),

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger, \quad (3)$$

and we assume that they act nontrivially on at most a fixed number d of qubits at a time. That is, the Pauli decomposition of Kraus operators,

$$E_k = \sum_{\vec{\alpha} \in I_d} e_{k\vec{\alpha}} \sigma_{\vec{\alpha}}, \quad (4)$$

only takes nonzero values in the set $I_d := \{\vec{\alpha} \in \{0, x, y, z\}^n \mid \sigma_{\vec{\alpha}} \text{ acts nontrivially on at most } d \text{ qubits}\}$, where with σ_0 we mean the identity operator. This set has cardinality $D = \sum_{i=0}^d 3^d \binom{n}{i}$ and is therefore polynomial in n . A convenient way to parametrize channels of this kind is through the *process matrix* representation,

$$\mathcal{E}(\rho) = \sum_{\vec{\alpha}, \vec{\beta} \in I_d} \chi_{\vec{\alpha}\vec{\beta}} \sigma_{\vec{\alpha}} \rho \sigma_{\vec{\beta}}, \quad (5)$$

where we have made use of the expansion in Eq. (4) and defined as $\chi_{\vec{\alpha}\vec{\beta}} = \sum_k e_{\vec{\beta}k}^* e_{k\vec{\alpha}}$. While being exponentially large in the general case, channels following our ansatz are represented by process matrices of polynomial size: $\chi_{\vec{\alpha}\vec{\beta}} = (D \times D)$. They capture the intuition of stochastic errors occurring at once only on few qubits and are often referred to as d -degree (or *low-degree*) Pauli channels, a class that has already been explored in the literature [29]. To keep the model broadly applicable, we make no assumptions about qubit connectivity, allowing it to apply to quantum devices of arbitrary layout. In Appendix A we show that this ansatz well represents the error channels in two physically motivated settings. First, we consider weak single-qubit errors and show that the noise channel can be well approximated by a low-degree expansion provided the error probability scales at most as $\mathcal{O}(\sqrt{d/n - \log(1/\epsilon)/n})$, where ϵ denotes

the target precision of the reconstruction protocol. Second, we analyze weak spurious two-qubit interactions and find that the ansatz remains accurate when the interaction strength is at most $\mathcal{O}(\epsilon^{1/d}/n^2)$. These examples support the idea that low-degree channels capture realistic error models in regimes of practical interest (e.g., weak noise).

Importantly we notice that the error component $\mathcal{E}$ is a low-degree quantum channel, while the whole noisy implementation $\mathcal{C}$ might not be. For instance, if we consider a layer of Pauli gates, it will be represented by full weight Pauli strings and so will its noisy implementation. This observation motivates our work. With respect to previous techniques [29] we provide contributions that address two challenges:

- (1) first we define an efficient and experimentally friendly tomography scheme for low-degree quantum channels;
- (2) then we adapt the protocol to reconstruct the channel $\mathcal{E}$ while only having experimental access to $\mathcal{C}$.

We seek a tomography protocol that scales at most polynomially with system size and relies only on single-qubit state preparation and measurement. This requirement is consistent with our assumption that single-qubit operations are approximately noiseless, ensuring that the state preparation and measurement procedures do not introduce additional noise into the system.

The proposed tomography method is summarized in Fig. 1. By employing randomized Pauli state preparation and measurement, we show that the error channel can be learned to a desired accuracy in ℓ_2 -distance using only a logarithmic number of samples. In addition, we relate the diamond distance between two channels to the ℓ_1 -distance between their process matrices and show that the resulting reconstruction cost in diamond distance remains

polynomial in the system size. Furthermore, we demonstrate that the noisy component of the implementation [Eq. (2)] in many relevant cases can be extracted entirely in classical postprocessing—avoiding the need to invert the unitary layer on hardware, which would itself be an additional source of error. This inversion comes with an overhead at most polynomial in sample complexity. In what follows, we first introduce the tomography protocol and then adapt it to include the classical postprocessing part.

III. TOMOGRAPHY PROTOCOL

We consider a randomized Pauli measurement tomography scheme, which has been widely successful in the context of state tomography [31]. For each qubit, we sample randomly an initial Pauli eigenstate and a measurement basis from the Pauli basis. We denote the input and output states by $|\vec{s}\rangle$ and $|\vec{r}\rangle$, taking the six possible values $\{|+\rangle, |-\rangle, |i\rangle, |-i\rangle, |0\rangle, |1\rangle\}$. The probability $p(\vec{r}, \vec{s})$ of measuring an output $|\vec{r}\rangle$ given an input $|\vec{s}\rangle$ can be written as

$$\begin{aligned} p(\vec{r}, \vec{s}) &= \frac{1}{18^n} \langle \vec{r} | \mathcal{E}(|\vec{s}\rangle\langle \vec{s}|) | \vec{r} \rangle \\ &= \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}} F_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}), \end{aligned} \quad (6)$$

by introducing the function

$$F_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}) = \frac{1}{18^n} \langle \vec{r} | \sigma_{\vec{\alpha}} |\vec{s}\rangle \langle \vec{s}| \sigma_{\vec{\beta}} | \vec{r} \rangle, \quad (7)$$

with the $1/18^n$ prefactor chosen to ensure the overall normalization $\sum_{\vec{r}, \vec{s}} p(\vec{r}, \vec{s}) = 1$. We can interpret F as a matrix whose rows are indexed by the process matrix entries $\vec{\alpha}, \vec{\beta}$ and whose columns are indexed by the possible input and

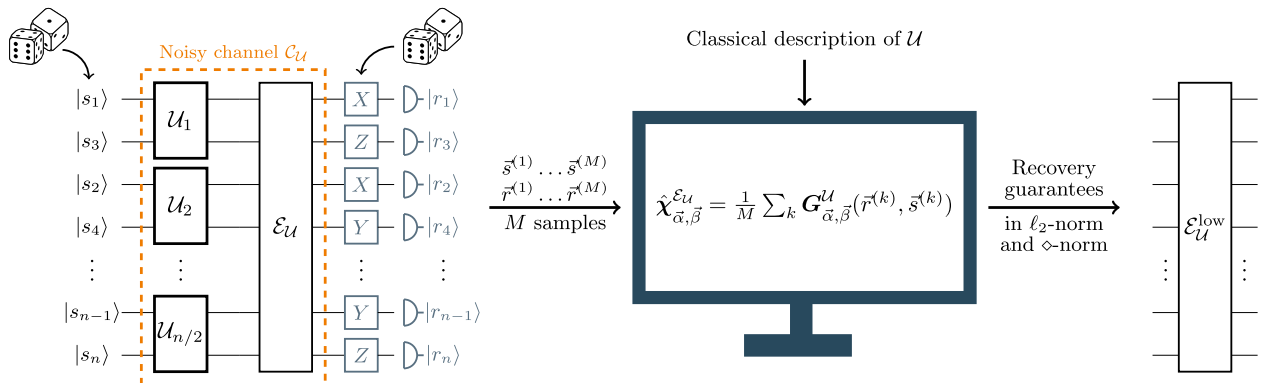


FIG. 1. Schematic summary of the proposed tomography protocol. A random set of single-qubit Pauli eigenstates is initialized, evolved through the noisy channel $\mathcal{C}_U$ and measured in a random Pauli basis. The classical outcome of the measurements is then sent to a classical computer that uses them alongside the classical description of U (e.g., its matrix representation in the computational basis) to compute the estimated process matrix $\chi^{\mathcal{E}_U}$ by averaging over the function G^U which converges entrywise to $\chi^{\mathcal{E}_U}$. The final output of the protocol is then the low-degree approximation of the noisy part of the implementation $\mathcal{E}_U$.

output states $\vec{r}, \vec{s}$. Following this formalism, we can regard both $\chi_{\vec{\alpha}\vec{\beta}}$ and $p(\vec{r}, \vec{s})$ as vectors indexed by $\vec{\alpha}, \vec{\beta}$ and $\vec{r}, \vec{s}$, respectively. In this notation, Eq. (6) can be rewritten as a matrix–vector product,

$$\vec{p} = \mathbf{F} \vec{\chi}. \quad (8)$$

To recover $\vec{\chi}$ from $\vec{p}$, we need to define a left inverse of $\mathbf{F}$, which we denote by $\mathbf{G}$. Any matrix $\mathbf{G}$ satisfying this condition allows us to write

$$\vec{\chi} = \mathbf{G} \vec{p}, \quad (9)$$

or, more explicitly,

$$\chi_{\vec{\alpha}\vec{\beta}} = \sum_{\vec{r}, \vec{s}} p(\vec{r}, \vec{s}) G_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}) = \mathbb{E}[G_{\vec{\alpha}\vec{\beta}}]. \quad (10)$$

Hence, computing the value $G_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s})$ for each snapshot allows one to recover the corresponding entry $\chi_{\vec{\alpha}\vec{\beta}}$ by averaging over sufficiently many samples. An example of such a function is provided in the well-known shadow-tomography protocol [20,31], where a specific choice of inverse is made. In particular, the inverse is taken to be $\mathbf{G} := (\mathbf{F}^\dagger \mathbf{F})^{-1} \mathbf{F}^\dagger$. We will later provide more details on this choice in the context of the shadow-tomography protocol.

The convergence to the actual value of $\chi_{\vec{\alpha}\vec{\beta}}$ will scale with the variance of $G_{\vec{\alpha}\vec{\beta}}$ divided by the total number of measurements. Therefore, the key ingredient to obtain an efficient tomography protocol is to control the scaling of the variance of $G_{\vec{\alpha}\vec{\beta}}$ with the system size [20,31]. In the following we show how to derive a choice of $\mathbf{G}$ whose variance remains constant with n for every entry of χ , if evaluated on snapshots coming from low-degree channels of the form (5).

We make use of the fact that, for product input and output states, the function $\mathbf{F}$ takes a product form

$$F_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}) = \prod_{i=1}^n f_{\alpha_i \beta_i}(r_i, s_i), \quad (11)$$

where

$$f_{\alpha_i \beta_i}(r_i, s_i) = \frac{1}{18} \langle r_i | \sigma_{\alpha_i} | s_i \rangle \langle s_i | \sigma_{\beta_i} | r_i \rangle, \quad (12)$$

and index i refers to the i th qubit. We restrict to functions taking the same product structure $\mathbf{G}_{\vec{\alpha}\vec{\beta}} = \prod_i g_{\alpha_i \beta_i}$, where $g_{\alpha_i \beta_i} = (f_{\alpha_i \beta_i})^{-1}$. More specifically we ask for

$$\sum_{r_i, s_i} g_{\gamma_i \delta_i}(r_i, s_i) f_{\alpha_i \beta_i}(r_i, s_i) = \delta_{\alpha_i \gamma_i} \delta_{\beta_i \delta_i}. \quad (13)$$

A well-known example of such a construction, as mentioned previously, is provided by the shadow-tomography

protocol [20,31]. In that case the single-qubit function $g_{\alpha_i \beta_i}(r_i, s_i)$ is given by

$$g_{\gamma_i \delta_i}^{\text{sh}}(r_i, s_i) = \text{Tr}[|\phi_{\gamma_i}\rangle \langle \phi_{\delta_i}| D(r_i, s_i)], \quad (14)$$

with $|\phi_{\gamma_i}\rangle = (\sigma_{\gamma_i} \otimes \mathbb{1}) |\phi^+\rangle$, $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and $D(r_i, s_i) = \mathcal{D}_{1/3}^{-1}(|r_i\rangle \langle r_i|) \otimes \mathcal{D}_{1/3}^{-1}(|s_i\rangle \langle s_i|)$, where $\mathcal{D}_{1/3}^{-1}(A) = 3A - \mathbb{1}$. As already noted in previous work, the application of this function to process tomography often leads to variance that grows exponentially with system size [20]. In Appendix B we show that this also applies to the estimation of the entries of the process matrix. In contrast, our construction in the next section provides a bounded-variance alternative tailored to low-degree channels.

Building on this observation, we note that—as already discussed in Refs. [32–34]—the choice of the inverse map $D(r_i, s_i)$ is not unique, and different choices can lead to exponentially different variances [32]. We will exploit this freedom to identify a form of $\mathbf{G}$ that minimizes the variance specifically for low-weight error channels.

We use our formalism to cast the problem into a quadratic minimization. Let us think of the single-qubit function $\mathbf{f} := f_{\alpha_i \beta_i}(r_i, s_i)$ as a rectangular matrix of dimension (36×16) with rows indexed by r_i, s_i and columns indexed by α_i, β_i . Given the dimension count there will be at least 20 vectors in the left kernel of $\mathbf{f}$. We denote $\mathbf{K}$ the matrix such that $K_j(r_i, s_i)$ is the j th element of such vectors. For every choice of $\mathbf{g}$ then we have that $\mathbf{g} + \vec{x}^T \mathbf{K}$ is also a valid choice of inverse, with $\vec{x}$ being a vector of 20 free parameters.

We can now choose $\vec{x}$ to get the minimum variance of the function $\mathbf{G}$, expressed as

$$\begin{aligned} \text{Var}[G_{\vec{\gamma}\vec{\delta}}] &= \mathbb{E}[|G_{\vec{\gamma}\vec{\delta}}|^2] - |\mathbb{E}[G_{\vec{\gamma}\vec{\delta}}]|^2 \\ &= \sum_{\substack{\vec{r}, \vec{s} \\ \vec{\alpha}, \vec{\beta}}} \chi_{\vec{\alpha}\vec{\beta}} |G_{\vec{\gamma}\vec{\delta}}(\vec{r}, \vec{s})|^2 F_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}) - |\chi_{\vec{\gamma}\vec{\delta}}|^2 \\ &\leq \sum_{\substack{\vec{r}, \vec{s} \\ \vec{\alpha}, \vec{\beta}}} \chi_{\vec{\alpha}\vec{\beta}} |G_{\vec{\gamma}\vec{\delta}}(\vec{r}, \vec{s})|^2 F_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}). \end{aligned} \quad (15)$$

Making use of the single-qubit structure of the problem allows to write Eq. (15) as

$$\sum_{\vec{\alpha}, \vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}} \prod_{i=1}^n \left(\sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \right). \quad (16)$$

We now exploit the d -Pauli-weight channel assumption. Since $\alpha_i, \beta_i, \gamma_i$ or δ_i can differ from the identity on a fixed number of qubits (given by $4d$), we can split the above

product in the following way:

$$\prod_{i=1}^{n-4d} \left(\sum_{r_i, s_i} |g_{00}(r_i, s_i)|^2 f_{00}(r_i, s_i) \right) \times \prod_{j=1}^{4d} \left(\sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \right), \quad (17)$$

where we have assumed without loss of generality that on the first $n - 4d$ qubits $\alpha_i, \beta_i, \gamma_i$ or δ_i are the identity. We can see that in order for this product (and therefore the variance) to not scale exponentially in system size it is sufficient to ask for

$$\sum_{r_i, s_i} |g_{00}(r_i, s_i)|^2 f_{00}(r_i, s_i) \leq 1. \quad (18)$$

Enforcing this condition translates into the following quadratic minimization problem:

$$\min_{\vec{x}} \left\{ \sum_{r_i, s_i} |g_{00}^{\text{sh}} + \vec{x}^T \mathbf{K}|^2 (r_i, s_i) f_{00}(r_i, s_i) \right\}, \quad (19)$$

where we have fixed the initial choice of $\mathbf{g}$ to be the one in Eq. (14). Quadratic problems admit a well-known analytic solution, leading in this case to

$$g_{00}^{\min}(r_i, s_i) = \begin{cases} -\frac{7}{2} & \text{for } r_i = \bar{s}_i, \\ 1 & \text{otherwise,} \end{cases} \quad (20)$$

where with $r_i = \bar{s}_i$ we mean that r_i and s_i are in the same Pauli basis, but different eigenstates (e.g., $|r_i\rangle = |+\rangle$ and $|s_i\rangle = |-\rangle$). By direct inspection one can see that $g_{00}^{\min}$ saturates the inequality in Eq. (18). To obtain a better scaling of the sample complexity it is reasonable to apply the same minimization procedure defined in Eq. (19) to all the entries γ_i, δ_i ,

$$\min_{\vec{x}} \left\{ \sum_{r_i, s_i} |g_{\gamma_i \delta_i}^{\text{sh}} + \vec{x}^T \mathbf{K}|^2 (r_i, s_i) f_{00}(r_i, s_i) \right\}. \quad (21)$$

We then define $\mathbf{g}^{\min}$ as the matrix that results from the above minimization of all entries. We refer to Appendix C for an explicit expression of such function. Now, using $\mathbf{g}^{\min}$, we can prove that

$$\text{Var}[G_{\vec{\gamma} \vec{\delta}}^{\min}] \leq C^{4d}, \quad (22)$$

with C being the smallest constant such that $\sum_{r_i, s_i} |g_{\gamma_i \delta_i}^{\min}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \leq C$ for every $\alpha_i, \beta_i, \gamma_i, \delta_i$ and $\mathbf{G}^{\min}$ being the product of $\mathbf{g}^{\min}$ over all the qubits. This last equation implies that, making use of the median of means

(MoM) estimator [35], the sample complexity M to estimate with probability $1 - \delta$ one entry of χ with precision ϵ is

$$M = \mathcal{O} \left(\frac{e^d}{\epsilon^d} \log \left(\frac{1}{\delta} \right) \right), \quad (23)$$

which is constant in system size. If instead we wish to estimate the ℓ_2 -distance of the channel resulting from the estimated $\chi_{\vec{\alpha} \vec{\beta}}$ and the actual noise component $\mathcal{E}$ we need

$$M = e^{\mathcal{O}(d \log(1/\epsilon))} \cdot \mathcal{O} \left(\log \left(\frac{n}{\delta} \right) \right), \quad (24)$$

which is now logarithmic in system size. Finally the scaling for the reconstruction in diamond distance is

$$M = n^{\mathcal{O}(d)} e^{\mathcal{O}(\log(1/\epsilon))} \cdot \mathcal{O} \left(\log \left(\frac{1}{\delta} \right) \right). \quad (25)$$

It is interesting to note that, to obtain the bounds on the variance, it is sufficient to assume that the process is trace nonincreasing (as shown in Lemma C2). Consequently, our protocol can be readily applied to trace nonincreasing channels, which describe, for instance, physical processes such as leakage [36].

From now on when talking about $\mathbf{g}$ we will mean the function resulting from the minimization procedure: $\mathbf{g}^{\min}$.

IV. CLASSICAL POSTPROCESSING

The tomography protocol described above is efficient in the case of low-degree channels. Yet, what one has direct experimental access to is the noisy implementation $\mathcal{C}$, which includes the unitary-layer contribution. Even in the case of low noise, such a map is far from being a low-degree channel, while the noisy part $\mathcal{E} = \mathcal{C} \circ \mathcal{U}^\dagger$ is. However, implementing $\mathcal{U}^\dagger$ in the quantum device would create a new source of error and would not allow for the correct estimate of the noisy part $\mathcal{E}$.

We thus seek to implement the inverse gate directly in classical postprocessing, essentially by adapting the function $\mathbf{G}$ to undo the action of $\mathcal{U}$. This process is not guaranteed to preserve the efficiency of the tomography protocol. Yet, we find that the protocol remains scalable (with just a polynomial overhead) for the case of layers containing a polylogarithmic number of non-Clifford gates.

Let us explain the postprocessing protocol in detail: first, recall that we want to reconstruct the process matrix of the noise component from Eq. (10), namely,

$$\chi_{\vec{\alpha} \vec{\beta}}^{\mathcal{E}} = \frac{1}{18^n} \sum_{\vec{r}, \vec{s}} \langle \vec{r} | \mathcal{C} \circ \mathcal{U}^\dagger(|\vec{s}\rangle \langle \vec{s}|) | \vec{r} \rangle G_{\vec{\alpha} \vec{\beta}}(\vec{r}, \vec{s}). \quad (26)$$

However, the randomized measurement data is sampled according to the noisy implementation of $\mathcal{U}$, that is,

$$p_{\mathcal{C}}(\vec{r}, \vec{s}) = \frac{1}{18^n} \langle \vec{r} | \mathcal{C}(|\vec{s}\rangle\langle\vec{s}|) | \vec{r} \rangle. \quad (27)$$

With simple manipulation we can rewrite

$$\begin{aligned} \chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}} &= \frac{1}{18^n} \sum_{\vec{r}, \vec{s}'} \langle \vec{r} | \mathcal{C}(|\vec{s}'\rangle\langle\vec{s}'|) | \vec{r} \rangle G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s}') \\ &= \mathbb{E}_{\sim \mathcal{C}}[G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s}')], \end{aligned} \quad (28)$$

where we use the subscript $\sim \mathcal{C}$ to indicate that the expectation value is taken with respect to the probability distribution generated by $\mathcal{C}$. Also, we have explicated the action of $\mathcal{U}^\dagger$ on its components: $\mathcal{U}^\dagger(|\vec{s}\rangle\langle\vec{s}|) = \sum_{\vec{s}'} u_{\vec{s}, \vec{s}'} |\vec{s}'\rangle\langle\vec{s}'|$ and defined the rotated $\mathbf{G}$ function,

$$G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s}') = \sum_{\vec{s}} u_{\vec{s}, \vec{s}'} G_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}). \quad (29)$$

In order to have efficient sample complexity we need the function above to have variance bounded by $\text{poly}(n)$ on the distribution (27),

$$\text{Var}[G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}] = \mathbb{E}_{\sim \mathcal{C}}[|G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}|^2] - |\mathbb{E}_{\sim \mathcal{C}}[G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}]|^2 \quad (30)$$

$$\leq \frac{1}{18^n} \sum_{\vec{r}, \vec{s}} |G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s})|^2 \langle \vec{r} | \mathcal{C}(|\vec{s}\rangle\langle\vec{s}|) | \vec{r} \rangle. \quad (31)$$

We now invert Eq. (2) to write $\mathcal{C} = \mathcal{E} \circ \mathcal{U}$, that is, the noisy implementation can be modeled by the noiseless $\mathcal{U}$, followed by the error channel. Then, Eq. (31) becomes

$$\frac{1}{18^n} \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}} \sum_{\vec{r}, \vec{s}} |G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s})|^2 \langle \vec{r} | \sigma_{\vec{\alpha}} \mathcal{U}(|\vec{s}\rangle\langle\vec{s}|) \sigma_{\vec{\beta}} | \vec{r} \rangle. \quad (32)$$

Let us now estimate the scaling of this term for a few interesting scenarios.

Unitary on polylogarithmic number of qubits. Let us consider a unitary map $\mathcal{U}$ that only acts on a set Q of qubits of cardinality q . Then, we can write

$$G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s}) = G_{\vec{\gamma}\vec{\delta}}^Q \left(\prod_{i \in Q} g_{\gamma_i \delta_i}(r_i, s_i) \right), \quad (33)$$

where $G_{\vec{\gamma}\vec{\delta}}^Q$ is the part of the function supported on Q . It follows that $\chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{C}}$ will differ from $\chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}}$ on at most q qubits for every $\vec{\alpha}, \vec{\beta}$, having nonzero entries for Pauli strings

that have weight at most $2d + q$. This leads to a favorable sample complexity, with variance scaling as

$$\text{Var}(G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s})) \leq K^q C^{Ad}, \quad (34)$$

where K is a constant. Therefore by taking $q = \mathcal{O}(\text{poly} \log(n))$ we can see that, accepting a polynomial overhead in n , the sample complexity remains efficient,

$$M = \mathcal{O} \left(n^p \frac{e^d}{\epsilon} \log \left(\frac{1}{\delta} \right) \right), \quad (35)$$

where p is the product of the exponents appearing in the polynomial dependence on $\log$ of the support of such gates. For a detailed derivation we refer to Appendix C 1.

Layer of two-qubit gates. Take a layer of $n/2$ unitaries $\mathcal{U}_{\mathbf{j}}$, acting each on a pair of qubits $\mathbf{j} = (j_0, j_1)$.

$$\mathcal{U} = \mathcal{U}_0 \otimes \mathcal{U}_1 \otimes \cdots \mathcal{U}_{n/2}, \quad (36)$$

with no restriction on the connectivity of the device. By exploiting the tensor product structure, $G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}$ becomes a product of $n/2$ two-qubit functions,

$$G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s}) = \sum_{\vec{s}'} u_{\vec{s}, \vec{s}'} G_{\vec{\alpha}\vec{\beta}}(\vec{r}, \vec{s}') \quad (37)$$

$$= \prod_{\mathbf{j} \in J} g_{\vec{\alpha}\vec{\beta}\vec{\gamma}}^{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}), \quad (38)$$

where we defined the functions $g_{\vec{\alpha}\vec{\beta}\vec{\gamma}}^{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}) = \sum_{\vec{s}'_{\mathbf{j}}} u_{\vec{s}_{\mathbf{j}}, \vec{s}'_{\mathbf{j}}} g_{\alpha_{j_0} \beta_{j_0}}(r_{j_0}, s'_{j_0}) g_{\alpha_{j_1} \beta_{j_1}}(r_{j_1}, s'_{j_1})$ acting on the pair of qubits (j_0, j_1) . Since we assume $\mathcal{E}$ to be a few-body channel, the relevant contribution to Eq. (32) comes for terms with $\alpha_i, \beta_i, \gamma_i$ and δ_i equal to the identity. In fact, given the above product structure of $G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s})$, we can split the contributions similarly to what done in Eq. (17). Then we can see that the exponential scaling can be avoided if the terms corresponding to all indexes equal to identity are smaller or equal to 1. In Appendix C 2 we show that bounding such contributions results in the following minimization problem:

$$\min_{\vec{x}} \sum_{\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}} |g_{00}^{\mathcal{U}_{\mathbf{j}}} + \vec{x}^T \mathbf{K}|^2 p_{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}), \quad (39)$$

where $p_{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}) = \frac{1}{18^2} \langle \vec{r}_{\mathbf{j}} | \mathcal{U}_{\mathbf{j}}(|\vec{s}_{\mathbf{j}}\rangle\langle\vec{s}_{\mathbf{j}}|) | \vec{r}_{\mathbf{j}} \rangle$. Notice that now, since we are considering a two-qubit space, $\vec{x}$ encodes 1040 free parameters.

Solving the corresponding quadratic problem allows us to find that Clifford gates always correspond to a minimum of 1. In Appendix C 3 we show that this leads to the

variance scaling as

$$\text{Var}(G_{\vec{\gamma}\vec{\delta}}) \leq nC^{4d}. \quad (40)$$

That is, one can estimate with probability $1 - \delta$ one entry of χ with precision ϵ via the relation

$$\hat{\chi}_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}} = \frac{1}{M} \sum_{k=1}^M \mathbf{G}_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}^{(k)}, \vec{s}^{(k)}), \quad (41)$$

with sample complexity scaling as

$$M = \mathcal{O}\left(n \frac{e^d}{\epsilon} \log\left(\frac{1}{\delta}\right)\right). \quad (42)$$

Notice that now the sample complexity required to estimate the channel in ℓ_2 -norm is multiplied by a linear factor in system size,

$$M = e^{\mathcal{O}(d \log(1/\epsilon))} \cdot \mathcal{O}\left(n \log\left(\frac{n}{\delta}\right)\right). \quad (43)$$

The argument carries over to the diamond distance as well; however, because the scaling without postprocessing already contains a factor of $n^{\mathcal{O}(d)}$, the scaling for reconstruction in diamond norm remains unchanged. Despite the bound being polynomial in system size, one can find numerical evidence that in many relevant cases the scaling of Eq. (42) is actually constant, similar to what found for the case without postprocessing (see Appendix C4 for more details). We also confirmed numerically that the postprocessing scheme for non-Clifford gates leads to an exponential sample complexity (see Appendix C4). The code to generate $\mathbf{g}^{\mathcal{U}_i}$ for any one or two-qubit unitary is publicly available in Ref. [37], alongside the precomputed functions for CNOT and ISWAP gates.

Combining together the discussion of the two subcases, we find that our postprocessing tomography protocol shows at most polynomial scaling for any layer of two-qubit gates with a polylogarithmic number of non-Clifford gates. In fact, for the reconstruction in ℓ_2 -norm, we get a final sample complexity of

$$M = e^{\mathcal{O}(d \log(1/\epsilon))} \cdot \mathcal{O}\left(n^{p+1} \log\left(\frac{n}{\delta}\right)\right). \quad (44)$$

Finally the scaling in diamond distance reads

$$M = n^{\mathcal{O}(d)+p} e^{\mathcal{O}(\log(1/\epsilon))} \cdot \mathcal{O}\left(\log\left(\frac{1}{\delta}\right)\right). \quad (45)$$

V. NUMERICAL ANALYSIS

To showcase the scalability our method, we simulated numerically the action of a known error channel and reconstructed its low-degree process matrix by means of our

tomography protocol. We focus our numerically analysis on two key aspects: first we show the applicability of our model by computing numerically the variance of $\mathbf{G}$ for a correlated, coherent, low-degree error channel. Then we simulate the sampling procedure in the case of a single-qubit tensor product error channel.

For the first task we chose an error channel $\mathcal{E}(\rho) = \sum_{k=1}^n E_k \rho E_k^\dagger$ with Kraus operators of the form

$$E_k = p \left(\mathbb{1} + i\epsilon \sum_{m=1}^n M_{km} \sigma_x^{(m)} \right), \quad (46)$$

where with $\sigma_x^{(m)}$ we mean the Pauli σ_x acting on qubit m . We can enforce the normalization $\sum_k E_k^\dagger E_k = \mathbb{1}$ by choosing $\mathbf{M}$ to be a normal matrix and $p = 1/(n(1 + \epsilon^2))$. We make the explicit choice $\mathbf{M} = \exp(-2/n\mathbf{A})$ with $\mathbf{A}$ being an anti-symmetric, tridiagonal matrix with periodic boundary conditions. This choice encapsulates the idea of correlated, coherent errors with magnitude decaying exponentially over an interaction length of $n/2$. Results are reported in Fig. 2 for the entries of $\mathbf{G}$ corresponding to $(\alpha_0, \beta_0) = (0, 0)$ and $(0, x)$ and 0 otherwise. The numerical analysis shows that the variance saturates to a constant.

For the simulation of the sampling procedure, we chose a 1D chain of qubits and the error channel as a tensor product of single-qubit channels of the form

$$\begin{aligned} \mathcal{E}_l(\cdot) = & p_l \sigma_z e^{-i\gamma_l \sigma_x}(\cdot) e^{i\gamma_l \sigma_x} \sigma_z \\ & + (1 - p_l) e^{-i\gamma_l \sigma_x}(\cdot) e^{i\gamma_l \sigma_x}, \end{aligned} \quad (47)$$

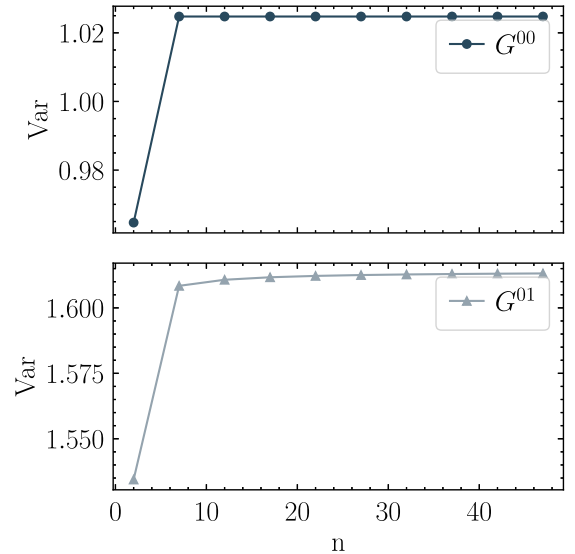


FIG. 2. Numerical computation of the variance of some entries of $\mathbf{G}$ for an error channel with Kraus operators of the form (46). The computation is carried for the entries corresponding to $\vec{\alpha} = \vec{\beta} = (0, \dots, 0)$ (top panel) and $\vec{\alpha} = (0, \dots, 0)$, $\vec{\beta} = (x, 0, \dots, 0)$ (bottom panel).

where the parameters p and γ decay exponentially with respect to the distance l of the qubit to the center,

$$p_l = p_0 \exp(-l), \quad \gamma_l = \gamma_0 \exp(-l). \quad (48)$$

We have chosen $p_0 = \gamma_0 = 0.1$. It is important to notice that while this channel does not strictly respect the low-degree ansatz it can be approximated by one, well enough for our protocol to be efficient. For a detailed analysis of the error committed in such approximation for similar cases refer to Appendix A. We run the simulation of the reconstruction of some entries of χ . More specifically we computed the number of samples needed to reach a fixed threshold of error ϵ for different system size, reaching up to 100 qubits.

We studied two scenarios, depicted in Figs. 4(a) and 4(b). First, the application of a single ISWAP gate in the center of the chain and then a full layer of nearest-neighbor

ISWAP gates. In the first case we do not make use of the rotated $\mathbf{G}$ function and therefore the average converges to the process matrix χ^c of the noisy implementation of the gate $\mathcal{C} = \mathcal{U}_{\text{ISWAP}} \circ \bigotimes_j \mathcal{E}_j$. We made this choice since the noisy implementation of one single ISWAP respects the ansatz of being low degree and therefore does not require the postprocessing procedure. In the second case, however, the contribution of the layer of gates generates a noisy channel $\mathcal{C}$ with non-negligible higher-order Pauli components. Therefore, we make use of the rotated $\mathbf{G}$ function defined in Eq. (29) to eliminate the action of the gates and optimized to minimize the variance [Eq. (39)]. We find that its average value converges correctly to the process matrix of just the error channel χ^ϵ .

We applied the procedure mentioned above for the reconstruction of three different entries of the process matrix. Figure 3 reports the results for different values of $\vec{\alpha}, \vec{\beta}$. In the top panel $\vec{\alpha} = \vec{\beta} = (0, \dots, 0)$ which

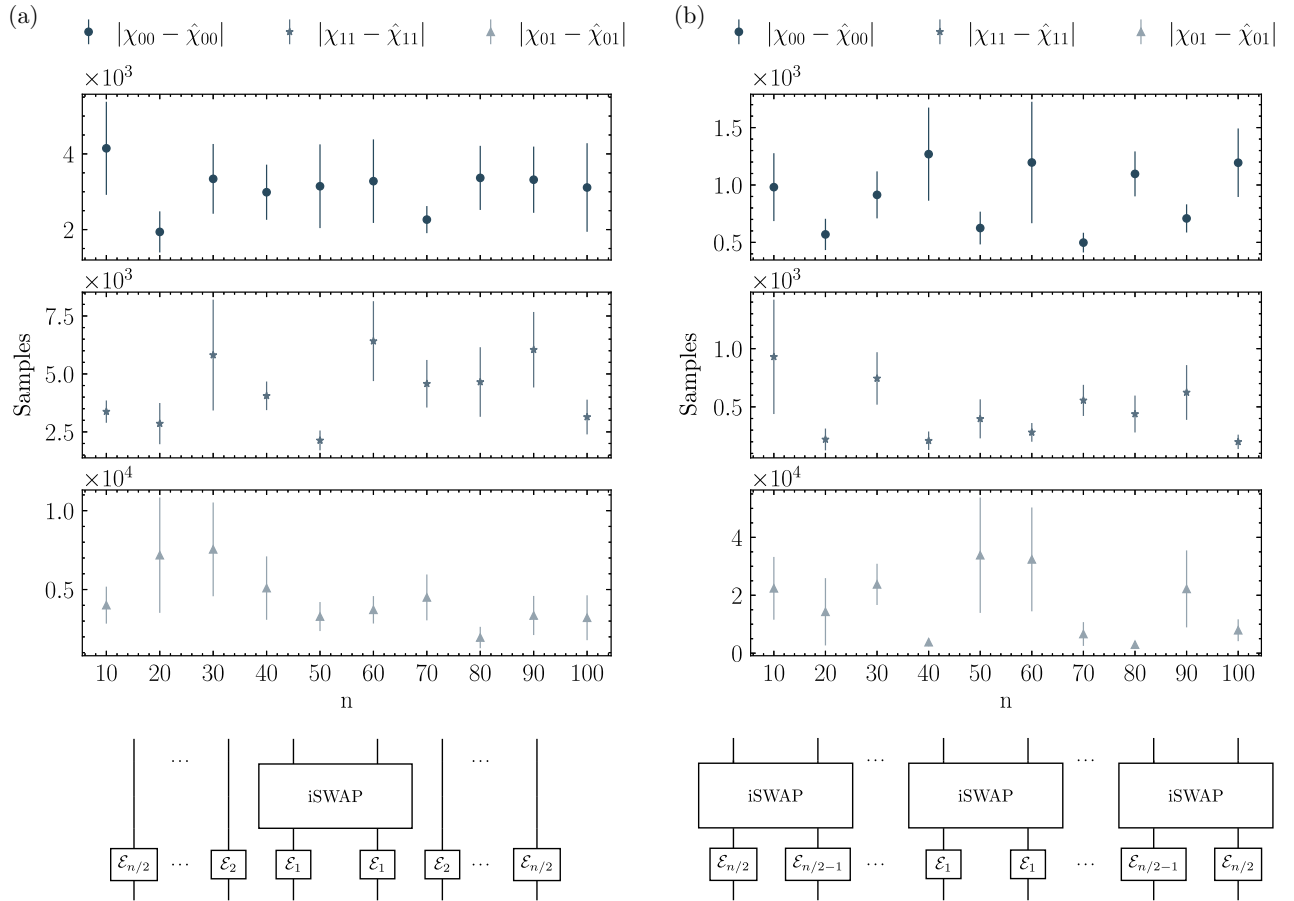


FIG. 3. Number of samples needed to reach a precision of 0.05 in the reconstruction of the entry corresponding to $\vec{\alpha} = (\alpha_0, \dots, 0)$ and $\vec{\beta} = (\beta_0, \dots, 0)$ for (from top to bottom) $(\alpha_0, \beta_0) = (0, 0), (1, 1), (0, 1)$. We report the results for two systems: (a) single ISWAP gate and (b) layer of ISWAP gates. We considered the convergence to be reached if for 500 consecutive samples $|\hat{\chi}_{\vec{\alpha}\vec{\beta}} - \chi_{\vec{\alpha}\vec{\beta}}| < 0.05$, where with $\hat{\chi}_{\vec{\alpha}\vec{\beta}}$ we mean the process matrix resulting from sampling. Every point in the plots is the result of an average over 10 independent samplings and the error bar shows the standard deviation from the mean. Below each plot is a representation of the corresponding circuit used in the numerics.

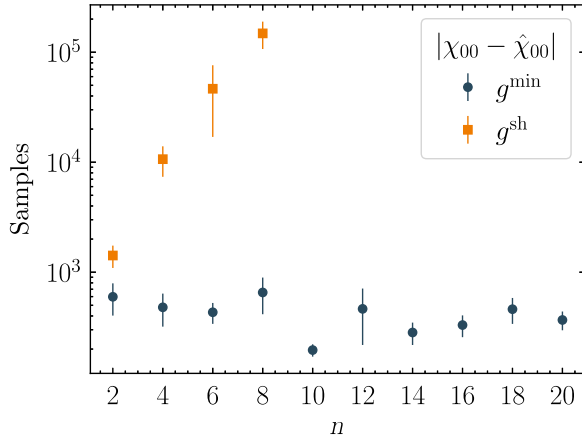


FIG. 4. Number of samples needed to reach a precision of 0.05 in the reconstruction of the entry corresponding to $\vec{\alpha} = \vec{\beta} = (0, \dots, 0)$ for a layer of ISWAP gates. We considered the convergence to be reached if for 100 consecutive samples $|\hat{\chi}_{\vec{\alpha}\vec{\beta}} - \chi_{\vec{\alpha}\vec{\beta}}| < 0.05$, where with $\hat{\chi}_{\vec{\alpha}\vec{\beta}}$ we mean the process matrix resulting from sampling. In blue the data obtained using the optimized function $\mathbf{g}^{\min}$ and in orange the data from using $\mathbf{g}^{\text{sh}}$. Every point in the plot is the result of an average over 10 independent samplings and the error bar shows the standard deviation from the mean.

corresponds to the all-identity case and is related to the fidelity of the process. On the central panel we estimate the entry corresponding to $\alpha_0, \beta_0 = x$ and all other entries equal to zero, which encodes the probability of having a bit flip error on the first qubit. We then turned to estimating an off-diagonal element and chose the entry corresponding to $\alpha_0 = 0, \beta_0 = x$ and all other entries equal to zero (bottom panel). Results reported show that for both cases we are able to reach the same error with a constant number of samples.

In Fig. 4 we show the difference in the scaling of the number of samples needed to reach an error of $\epsilon = 0.05$, using two different $\mathbf{g}$ functions. In blue we can see that using the optimized functions shows a constant scaling while in orange we see that the scaling is exponential if we simply use the function defined in Eq. (29) without minimizing the variance. While we are aware that shadow tomography is not the most efficient method to compute the fidelity of a process (i.e., χ_{00}), we report the comparison as a proof of principle that exponential advantage over classical shadow is possible.

VI. CONCLUSION

In this work, we demonstrated an efficient tomography protocol for quantum channels that conform to a low-degree ansatz, which corresponds to a perturbative regime in terms of the number of Pauli errors. This ansatz was first introduced in Ref. [29] and has since been employed in the analysis of various tomography methods [21,28,29].

Our protocol ensures logarithmic complexity in the system size and is experimentally practical, requiring only the preparation of product states in Pauli eigenbasis and measurements in the same basis. In particular, we provide a complete reconstruction of the channel in logarithmic sample complexity, with guarantees in the ℓ_2 -norm and diamond norm of the channel while only making use of single-qubit state preparation and measurement. We showcase how to apply the tomography protocol to characterize the noise affecting a layer of two-qubit gates in a quantum computer. A key advantage of our approach is that the inversion of the unitary gate layer, for many relevant cases, is handled entirely through classical postprocessing, avoiding the need to physically implement the inverse on quantum hardware. While only requiring at most polynomial overhead in sample complexity. This allows for a self-consistent tomography procedure.

A natural next step is to investigate how commonly used figures of merit—such as average gate fidelity and diamond norm error—translate into the process matrix representation within our ansatz. Understanding this relationship could yield deeper insight into the operational performance of the reconstructed channels. Another important avenue concerns the impact of SPAM errors on the protocol. Since these errors remain a dominant source of inaccuracy in current quantum platforms [38], studying how they affect our method—and developing strategies to mitigate them—will be essential for advancing the robustness and applicability of this framework in real-world experiments. In this regard, it may be possible to adapt techniques that have been successfully employed within the shadow-tomography formalism, such as robust shadow estimation [39] and probabilistic error cancellation [40,41], to our setting. These approaches could provide a path toward mitigating SPAM-induced biases without requiring significant changes to the protocol. However, the precise adaptation of these techniques to our protocol goes beyond the scope of the present work and is left for future investigation.

ACKNOWLEDGMENTS

M.C. thanks Patrick Emonts, Giacomo Giudice, Erickson Tjoa, and Dominik Wild (in alphabetical order) for fruitful discussions. F.B. thanks Daniel Malz for insightful discussion during the initial stages of the projects. We extend special thanks to Giacomo Giudice for valuable feedback on the manuscript. We thank the anonymous Referee B for insightful feedback that significantly improved this work, in particular by pointing out a way to reduce the scaling of the protocol from polynomial to logarithmic and to extend the postprocessing guarantee from a constant to a polylogarithmic number of non-Clifford gates. M.C. and I.C. acknowledge support from the German Federal Ministry of Education and Research (BMBF) via the project FermiQP (No. 13N15889). The work at MPQ is

partly funded by THEQUCO as part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. F.B. acknowledges financial support from the ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union—NextGenerationEU and from European Union’s Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Action for Project No. 101148556 (ENCHANT). The views and opinions expressed here are solely those of the authors and do not necessarily reflect those of the funding institutions. Neither of the funding institution can be held responsible for them.

DATA AVAILABILITY

The data that support the findings of this article are openly available [37].

APPENDIX A: PHYSICAL MODELS

In this section we will provide explicit examples in which our ansatz approximates well an error channel.

In general to compute the distance in ℓ_2 -norm of the channel to its truncated version one would have to sum the modulus square of the elements with degree bigger than d . In fact for Parseval’s identity for super operators we get [29]

$$\|\mathcal{E} - \mathcal{E}_d\|_2^2 = \sum_{\vec{\alpha}, \vec{\beta}} \left| \chi_{\vec{\alpha}\vec{\beta}} - \chi_{\vec{\alpha}\vec{\beta}}^d \right|^2 = \sum_{\vec{\alpha}, \vec{\beta} \notin I_d} \left| \chi_{\vec{\alpha}\vec{\beta}} \right|^2, \quad (\text{A1})$$

where we use the index d to indicate the truncation at degree d . Since χ is a positive semidefinite matrix this quantity can be bounded with the diagonal elements,

$$\|\mathcal{E} - \mathcal{E}_d\|_2^2 \leq \sum_{\vec{\alpha} \notin I_d} \chi_{\vec{\alpha}\vec{\alpha}} \chi_{\vec{\beta}\vec{\beta}}. \quad (\text{A2})$$

In the following we will compute this error in some interesting cases.

1. Small single-qubit errors

Assume that the noise acting on the quantum device can be represented as single-qubit noise acting on every qubit. For simplicity we will take the example of the a single-qubit bit flip, acting on qubit i with probability p ,

$$\mathcal{E}^{(i)}(\rho) = (1 - p)\rho + p\sigma_x^{(i)}\rho\sigma_x^{(i)}. \quad (\text{A3})$$

So the error channel acting on all n qubits will be

$$\mathcal{E}(\rho) = \bigotimes_{i=1}^n \mathcal{E}^{(i)}(\rho) = \sum_{k=0}^n (1 - p)^{n-k} p^k \sum_{i=0}^{\binom{n}{k}} P_i^{(k)} \rho P_i^{(k)}, \quad (\text{A4})$$

where $\{P_i^{(k)}\}_i$ are the k -degree Pauli x strings, which are $\binom{n}{k}$. For small p we can truncate the expression above up to a certain order. For instance if we choose to truncate to second order we will have

$$\begin{aligned} \mathcal{E}^{(i)}(\rho) &= \left(1 - np + \frac{n(n-1)}{2}p^2\right)\rho \\ &+ (p - (n-1)p^2) \sum_i \sigma_x^{(i)} \rho \sigma_x^{(i)} \\ &+ p^2 \sum_{i < j} \sigma_x^{(i)} \sigma_x^{(j)} \rho \sigma_x^{(j)} \sigma_x^{(i)} + \mathcal{O}(p^3), \end{aligned} \quad (\text{A5})$$

which respects the ansatz of Eq. (5).

To compute the error that we are making in the truncation we need to put together Eqs. (A1) and (A4).

$$\begin{aligned} \|\mathcal{E} - \mathcal{E}_d\|_2^2 &= \sum_{k=d+1}^n \binom{n}{k} [(1-p)^{n-k} p^k]^2 \\ &\leq \sum_{k=d+1}^n \binom{n}{k} (1-p^2)^{n-k} p^{2k} \\ &\leq \exp \left[-2n \left(\frac{d+1}{n} - p^2 \right) \right], \end{aligned} \quad (\text{A6})$$

where we have used that $(1-p)^{2(n-k)} \leq (1-p)^{n-k}$ for every $p \leq 1$ and then used Hoeffding’s bound on the tail of the binomial distribution. If we want the error to be bounded by ϵ we need

$$p < \left[\frac{d+1}{n} - \frac{1}{2n} \log \left(\frac{1}{\epsilon} \right) \right]^{1/2}. \quad (\text{A7})$$

Therefore, for small enough error probability p it is reasonable to expect low-degree Pauli channels to well represent the noise process.

2. Spurious interactions

Another common scenario is the presence of spurious interactions in a quantum device. In real life a quantum gate is implemented with an Hamiltonian evolution. Let us call H_0 the Hamiltonian that is implementing the desired gate. If such gate is a layer of two-qubit gates then H_0 will contain terms acting nontrivially on the couples that are affected,

$$H_0 = \sum_{\substack{\mathbf{j} \in J \\ p, q}} g_{pq}^{j_0 j_1} \sigma_p^{(j_0)} \sigma_q^{(j_1)}, \quad (\text{A8})$$

where $\{\mathbf{j}\}_J$ is the set that includes the couples (j_0, j_1) on which the two-qubit gates are acting and p, q are the

indexes of the Pauli matrices $\{0, x, y, z\}$. All the other possible interactions are instead represented by the following Hamiltonian:

$$H_{\text{int}} = \sum_{\substack{kl \\ pq}} c_{pq}^{kl} \sigma_p^{(k)} \sigma_q^{(l)}, \quad (\text{A9})$$

with k, l running possibly on all qubits present in the system. So H_{int} is a sum of local terms $h_{kl} = \sum_{pq} c_{pq}^{kl} \sigma_p^{(k)} \sigma_q^{(l)}$. When the gates corresponding to these are not being implemented this interaction should not affect the system. However it is often the case that some spurious interactions are left in the system. When trying to implement the layer of gates, the system will then evolve according to

$$\rho(t) = e^{-it(H_g + \alpha H_{\text{int}})} \rho e^{it(H_g + \alpha H_{\text{int}})}, \quad (\text{A10})$$

where α is a parameter that quantifies the amount of spurious interactions and is assumed to be small. Then the error channel defined in Eq. (2) in this case would be

$$\mathcal{E}(\rho) = e^{-it(H_g + \alpha H_{\text{int}})} e^{itH_g} \rho e^{-itH_g} e^{it(H_g + \alpha H_{\text{int}})}. \quad (\text{A11})$$

Now, we can write the derivative of the unitary evolution $U(t) = e^{-it(H_g + \alpha H_{\text{int}})} e^{itH_g}$ as

$$\frac{d}{dt} U(t) = e^{-it(H_g + \alpha H_{\text{int}})} (-i\alpha H_{\text{int}}) e^{itH_g} = -i\alpha U(t) \tilde{H}_{\text{int}}(t), \quad (\text{A12})$$

where we have defined $\tilde{H}_{\text{int}}(t) = e^{-itH_g} H_{\text{int}} e^{itH_g}$. Importantly we can notice that the Hamiltonian so defined only contains terms that act nontrivially on at most four qubits. We can see from Eq. (A12) that the unitary evolution can be written as

$$U(t) = T_R \exp \left[-i\alpha \int_0^t ds \tilde{H}_{\text{int}}(s) \right], \quad (\text{A13})$$

where with $T_R \exp$ we mean the inverse time ordered exponential. This can be expanded in powers of α as

$$\begin{aligned} U(t) &= \mathbb{1} + i\alpha \int_0^t ds \tilde{H}_{\text{int}}(s) \\ &\quad - \alpha^2 \int_0^t ds_2 \int_0^{s_2} ds_1 \tilde{H}_{\text{int}}(s_2) \tilde{H}_{\text{int}}(s_1) + \mathcal{O}(\alpha^3). \end{aligned} \quad (\text{A14})$$

The above truncated expression is again a low-degree evolution, where terms of order k in α correspond to up to $4k$ -weight operators. We can now call S_k the term of order

k in the above sum and write

$$\mathcal{E}(\rho) = \left(\sum_{j=0}^{\infty} S_j \right) \rho \left(\sum_{j'=0}^{\infty} S_{j'}^\dagger \right) = \sum_{m=0}^{\infty} \sum_{k=0}^m S_k \rho S_{m-k}^\dagger, \quad (\text{A15})$$

where we have grouped the terms of the same order in α . We can now write this in the process matrix representation, by expanding S_j in the Pauli basis,

$$\begin{aligned} \mathcal{E}(\rho) &= \sum_{m=0}^{\infty} \sum_{k=0}^m \left(\sum_{\vec{\alpha}} c_{\vec{\alpha}}^k P_{\vec{\alpha}} \right) \rho \left(\sum_{\vec{\beta}} (c_{\vec{\beta}}^{m-k})^* P_{\vec{\beta}} \right) \\ &= \sum_{\vec{\alpha}, \vec{\beta}} \sum_{m=0}^{\infty} \sum_{k=0}^m \chi_{\vec{\alpha}\vec{\beta}}^{km} P_{\vec{\alpha}} \rho P_{\vec{\beta}}, \end{aligned} \quad (\text{A16})$$

where we have defined $\chi_{\vec{\alpha}\vec{\beta}}^{km} = c_{\vec{\alpha}}^k (c_{\vec{\beta}}^{m-k})^*$. To bound the difference in ℓ_2 -norm it, from Eq. (A1) we can write

$$\|\mathcal{E} - \mathcal{E}_d\|_2 = \sum_{\vec{\alpha}, \vec{\beta} \notin I_d} \left| \sum_{m=0}^{\infty} \sum_{k=0}^m \chi_{\vec{\alpha}\vec{\beta}}^{km} \right|^2 \leq \sum_{\vec{\alpha}, \vec{\beta} \notin I_d} \sum_{m=0}^{\infty} \sum_{k=0}^m \left| \chi_{\vec{\alpha}\vec{\beta}}^{km} \right|^2. \quad (\text{A17})$$

We include in the sum all the terms for which $4m$ is bigger than d . In fact all the terms where one among S_k and S_{m-k} include contributions of order higher than $d/4$ are included in this set. Notice that those terms might contribute also to the lower degree part of the process matrix. Since those contributions are hard to isolate in general, we upper bound the sum by extending it to the set of all Paulis. Then we can write

$$\begin{aligned} \|\mathcal{E} - \mathcal{E}_d\|_2 &\leq \sum_{m=\lceil \frac{d}{4} \rceil}^{\infty} \sum_{k=0}^m \sum_{\vec{\alpha}, \vec{\beta}} \left| \chi_{\vec{\alpha}\vec{\beta}}^{km} \right|^2 \\ &= \sum_{m=\lceil \frac{d}{4} \rceil}^{\infty} \sum_{k=0}^m \sum_{\vec{\alpha}, \vec{\beta}} |c_{\vec{\alpha}}^k|^2 |(c_{\vec{\beta}}^{m-k})^*|^2 \\ &= \sum_{m=\lceil \frac{d}{4} \rceil}^{\infty} \sum_{k=0}^m \|S_k\|_2 \|S_{m-k}\|_2. \end{aligned} \quad (\text{A18})$$

Let us now bound the ℓ_2 -norm of S_k ,

$$\|S_k\|_2 = \alpha^k \frac{\|\tilde{H}_{\text{int}}\|_2^k t^k}{k!} \leq \frac{1}{k!} (\alpha h m t)^k \leq \frac{1}{k!} (\alpha h t (2n)^2)^k, \quad (\text{A19})$$

where we have used $\|\tilde{H}_{\text{int}}\|_2 = \|H_{\text{int}}\|_2 \leq \sum_{kl} \|h_{kl}\|_2$, then we bounded the norm of each local term with a constant h .

Finally we introduced m as the number of terms present in H_{int} which is smaller than $\binom{n}{2} \leq (2n)^2$ but could show smaller scaling (e.g., linear) for different device connectivity (e.g., nearest neighbors). Then from Eq. (A18) we get the bound

$$\begin{aligned} & \sum_{m=\lceil \frac{d}{4} \rceil}^{\infty} \sum_{k=0}^m \left(\frac{1}{(m-k)!k!} \right)^2 (\alpha h t (2n)^2)^{2m} \\ & \leq \sum_{m=\lceil \frac{d}{4} \rceil}^{\infty} (2\alpha h t (2n)^2)^{2m} \\ & = \frac{(2\alpha h t (2n)^2)^{2\lceil \frac{d}{4} \rceil}}{1 - (2\alpha h t (2n)^2)^2} < 2(2\alpha h t (2n)^2)^{2\lceil \frac{d}{4} \rceil}. \end{aligned} \quad (\text{A20})$$

Here we have used $\sum_{k=0}^m 1/(m-k)!k! \leq 2^m$ and the closed form solution of the geometric series, valid for $\alpha < 1/2ht(2n)^2$. Last inequality assumes that $2\alpha h t (2n)^2 < 1/2$. So to have error smaller than ϵ one would need

$$\alpha < \frac{1}{2ht(2n)^2} \left(\frac{\epsilon}{2} \right)^{1/(2\lceil \frac{d}{4} \rceil)}. \quad (\text{A21})$$

This condition is consistent with previous assumptions, provided that we can choose ϵ small enough. Therefore, as long as α is small enough with respect to system size, the error that we are making in truncating to a low-degree representation is negligible.

APPENDIX B: SHADOW-TOMOGRAPHY FUNCTION

An example of a function whose expectation values gives the desired observable [like $g_{\gamma_i \delta_i}(r_i, s_i)$ in our protocol (13)], is used in the well-known shadow-tomography protocol. In fact in such protocol one would link the quantity to estimate a certain observable of the reconstructed state. In our case, the state to reconstruct is the Choi state of channel $\mathcal{E}$,

$$\hat{\chi}_{\vec{\gamma} \vec{\delta}} = \text{Tr}[O_{\vec{\gamma} \vec{\delta}} \hat{\Lambda}_{\mathcal{E}}]. \quad (\text{B1})$$

It is easy to see that $\chi_{\vec{\gamma} \vec{\delta}} = \text{Tr}[(\sigma_{\vec{\gamma}} \otimes \mathbb{1}_n) |\phi^+\rangle \langle \phi^+|^{\otimes n} (\sigma_{\vec{\delta}} \otimes \mathbb{1}_n) \Lambda_{\mathcal{E}}]$, with $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. The reconstruction of the Choi state in shadow-tomography formalism is given by the average over the classical shadows of the input and output states as $\hat{\Lambda}_{\mathcal{E}} = \sum_{\vec{r}, \vec{s}} p(\vec{r}, \vec{s}) (\mathcal{D}_{1/3}^{\otimes -n}(|\vec{r}\rangle \langle \vec{r}|) \otimes \mathcal{D}_{1/3}^{\otimes -n}(|\vec{s}\rangle \langle \vec{s}|))$, where we have defined $\mathcal{D}_{\lambda}^{-1}$ as the inverse of the depolarizing channel with depolarizing probability λ . Putting it all together we have

that in shadow-tomography formalism,

$$\begin{aligned} \hat{\chi}_{\vec{\gamma} \vec{\delta}} &= \sum_{\vec{r}, \vec{s}} p(\vec{r}, \vec{s}) \text{Tr} \left[(\sigma_{\vec{\gamma}} \otimes \mathbb{1}_n) |\phi^+\rangle \langle \phi^+|^{\otimes n} (\sigma_{\vec{\delta}} \otimes \mathbb{1}_n) \right. \\ & \quad \left. \times (\mathcal{D}_{1/3}^{\otimes -n}(|\vec{r}\rangle \langle \vec{r}|) \otimes \mathcal{D}_{1/3}^{\otimes -n}(|\vec{s}\rangle \langle \vec{s}|)) \right], \end{aligned} \quad (\text{B2})$$

which translates to a single-qubit function that satisfies Eq. (13),

$$\begin{aligned} g_{\gamma_i \delta_i}^{\text{sh}}(r_i, s_i) &= \text{Tr} \left[(\sigma_{\gamma_i} \otimes \mathbb{1}) |\phi^+\rangle \langle \phi^+| (\sigma_{\delta_i} \otimes \mathbb{1}) \right. \\ & \quad \left. \times (\mathcal{D}_{1/3}^{-1}(|r_i\rangle \langle r_i|) \otimes \mathcal{D}_{1/3}^{-1}(|s_i\rangle \langle s_i|)) \right]. \end{aligned} \quad (\text{B3})$$

As mentioned in the main text, the function obtained by the product of Eq. (B3), despite being well known in literature, has variance growing exponentially with system size and for this reason cannot be used in practice. In fact we can provide a lower bound for its variance by making use of the following property of the function:

$$\begin{aligned} \sum_{r_i, s_i} |g_{\gamma_i \delta_i}^{\text{sh}}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) &= C_{\alpha_i \beta_i}^{\gamma_i \delta_i} \delta_{\alpha_i \beta_i} \quad \text{with} \\ C_{\alpha_i \beta_i}^{\gamma_i \delta_i} &\geq 11/8 \quad \forall \alpha_i, \beta_i, \gamma_i, \delta_i, \end{aligned} \quad (\text{B4})$$

where we recall that $f_{\alpha_i \beta_i}(r_i, s_i) = (1/18) \langle r_i | \sigma_{\alpha_i} | s_i \rangle \langle s_i | \sigma_{\beta_i} | r_i \rangle$. Then we can write the variance of $G_{\vec{\gamma} \vec{\delta}}^{\text{sh}}$ using $\mathbf{g}^{\text{sh}}$ as

$$\begin{aligned} \text{Var}(G_{\vec{\gamma} \vec{\delta}}^{\text{sh}}) &= \sum_{\vec{\alpha}, \vec{\beta}} \chi_{\vec{\alpha} \vec{\beta}} \prod_{i=1}^n \left(\sum_{r_i, s_i} |g_{\gamma_i \delta_i}^{\text{sh}}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \right) - |\chi_{\vec{\gamma} \vec{\delta}}|^2 \end{aligned} \quad (\text{B5})$$

$$\geq \left(\frac{11}{8} \right)^n \sum_{\vec{\alpha}, \vec{\beta}} \chi_{\vec{\alpha} \vec{\beta}} \delta_{\vec{\alpha} \vec{\beta}} - |\chi_{\vec{\gamma} \vec{\delta}}|^2 \quad (\text{B6})$$

$$\geq \left(\frac{11}{8} \right)^n - 1, \quad (\text{B7})$$

where we used the fact that the diagonal of χ sums to 1 and the fact that for every entry of χ we have $|\chi_{\gamma \delta}| \leq 1$. Notice that the above bound clearly scales exponentially with system size n .

APPENDIX C: PROOFS OF SAMPLE COMPLEXITY

In this section we will prove all the statements regarding sample complexity for our tomography protocol presented in Sec. III.

First let us prove under which conditions the variance of the function $\mathbf{G}$ defined in Eq. (10) remains constant with increasing system size. Such function can be expressed as a product of single-qubit functions that we call $\mathbf{g}$ defined in Eq. (13). There are two properties that we wish $\mathbf{g}$ to fulfill in order for the variance of $\mathbf{G}$ to be constant,

- (i) $\sum_{r_i, s_i} |g_{00}(r_i, s_i)|^2 f_{00}(r_i, s_i) \leq 1$,
- (ii) $\sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) = C_{\alpha_i \beta_i}^{\gamma_i \delta_i} \delta_{\alpha_i \beta_i} \forall \gamma_i, \delta_i$ for some constant factor $C_{\alpha_i \beta_i}^{\gamma_i \delta_i}$,

where $\mathbf{f}$ are the functions defined in Eq. (12). Let us now prove that these two conditions imply constant variance. The first thing to notice is that, given our ansatz, the number of elements for which $\alpha_i, \beta_i, \gamma_i$ or δ_i differ from the identity is constant and simply given by four times the Pauli weight of the problem: $4d$. So if C is the biggest constant such that $\sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \leq C$ for one

among $\alpha_i, \beta_i, \gamma_i$ or δ_i different from zero, we have that

$$\sum_{\vec{\alpha}, \vec{\beta} \in I_d} \chi_{\vec{\alpha} \vec{\beta}} \prod_{i=1}^n \left(\sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \right) \quad (C1)$$

$$\leq C^{4d} \sum_{\vec{\alpha}, \vec{\beta} \in I_d} \chi_{\vec{\alpha} \vec{\beta}} \delta_{\vec{\alpha} \vec{\beta}} \quad (C2)$$

$$= C^{4d} \text{Tr}[\chi] \leq C^{4d}, \quad (C3)$$

where in the first line we have specified the summation over the low-degree Paulis and used condition (ii). From line one to line two we have used the fact that $C_{00}^{00} \leq 1$ [condition (i)] and the limited number of elements in I_d that differ from zero. Lastly we used the fact that χ has trace equal or minor to one, due to trace nonincreasing property of the channel. Therefore if we were to find a function that satisfies conditions (i) and (ii) we would have

$$\text{Var}[G_{\vec{\alpha} \vec{\beta}}] \leq C^{4d}. \quad (C4)$$

For the sake of completeness let us report here the full description of the function $\mathbf{g}^{\min}$ obtained through the minimization procedure of Eq. (21), dropping the qubit index for clarity.

$$g_{00}^{\min}(r, s) = \begin{cases} -\frac{7}{2} & \text{for } r = \bar{s}, \\ 1 & \text{otherwise,} \end{cases} \quad (C5)$$

$$g_{\alpha\alpha}^{\min}(r, s) = \begin{cases} (-1)^{\sigma_{b(r)} \star \sigma_{\alpha}} \frac{9}{2} & \text{for } r = \bar{s}, \\ 0 & \text{otherwise,} \end{cases} \quad (C6)$$

$$g_{\alpha\beta}^{\min}(r, s) = \begin{cases} \left((-1)^{e(r)+e(s)} \frac{9}{4} \right) (\delta_{b(r)\alpha} \delta_{b(s)\beta} + \delta_{b(r)\beta} \delta_{b(s)\alpha}) & , \\ (-1)^{e(s)} \frac{9}{2} & \text{for } e(r) \neq e(s) \text{ and } b(r) = b(s) = [\sigma_{\alpha}, \sigma_{\beta}], \\ 0 & \text{otherwise,} \end{cases} \quad (C7)$$

$$g_{0\alpha}^{\min}(r, s) = (g_{\alpha 0}^{\min})^*(r, s) = \begin{cases} \varepsilon_{b(r)b(s)\alpha} (-1)^{e(r)+e(s)} \frac{9}{4} i & \text{for } b(r), b(s) \neq \alpha, \\ \left((-1)^{e(r)} \frac{9}{10} \delta_{b(r)\alpha} + (-1)^{e(s)} \frac{9}{10} \delta_{b(s)\alpha} \right) \left(\frac{1}{2} \right)^{\delta_{b(r)\alpha} \delta_{b(s)\alpha}} & \text{otherwise,} \end{cases} \quad (C8)$$

where we recall that with $r = \bar{s}$ we mean that r and s are in the Pauli same basis, but with opposite eigenstates. Then with $b(r)$ we indicate the basis of which r is eigenvector and with $e(r) \in \{0, 1\}$ the eigenvector. For instance $|-\rangle$ would have $b(-) = x$ and $e(-) = 1$. We have also introduced the operation $\star$ between two Pauli matrices that is so defined as

$$\sigma_{\alpha} \star \sigma_{\beta} = \begin{cases} 0 & [\sigma_{\alpha}, \sigma_{\beta}] = 0, \\ 1 & \text{otherwise.} \end{cases} \quad (C9)$$

Lastly with ε_{ijk} we mean the Levi-Civita tensor. Then we can see from direct computation that this function satisfies conditions (i) and (ii). We will now derive the exact expression of the sample complexity with respect to the variance of the function $\mathbf{G}^{\min}$, where we will drop the superscript.

Let us start from the well-known Chebyshev's inequality,

$$\Pr(|\hat{\chi}_{\vec{\alpha} \vec{\beta}} - \chi_{\vec{\alpha} \vec{\beta}}| > \epsilon) \leq \frac{\text{Var}(G_{\vec{\alpha} \vec{\beta}})}{M \epsilon^2}, \quad (C10)$$

where $\hat{\chi}_{\vec{\alpha}\vec{\beta}}$ is the estimator for $\chi_{\vec{\alpha}\vec{\beta}}$ and M is the total number of samples. In order to have that with probability $1 - \delta$ the estimated value does not differ from $\chi_{\vec{\alpha}\vec{\beta}}$ more than ϵ we need

$$M \sim \frac{\text{Var}(G_{\vec{\alpha}\vec{\beta}})}{\delta \epsilon^2}. \quad (\text{C11})$$

This expression, however, has a scaling of $1/\delta$ which is particularly bad. To avoid this we make use of the MoM estimator.

The MoM estimator works as follows: we divide our samples into K subsamples of dimension b , i.e., $M = Kb$. For each subsample we compute the mean of the samples, which leads to K estimators of $\hat{\chi}_{\vec{\alpha}\vec{\beta}}$,

$$\hat{\chi}_{\vec{\alpha}\vec{\beta}}^{(1)}, \dots, \hat{\chi}_{\vec{\alpha}\vec{\beta}}^{(K)}. \quad (\text{C12})$$

The median of means estimator is then taken to be the median of these estimators,

$$\hat{\chi}_{\vec{\alpha}\vec{\beta}}^{\text{MoM}} = \text{median}\{\hat{\chi}_{\vec{\alpha}\vec{\beta}}^{(1)}, \dots, \hat{\chi}_{\vec{\alpha}\vec{\beta}}^{(K)}\}. \quad (\text{C13})$$

If we now fix the size of the samples to be $b = 34\text{Var}(G_{\vec{\alpha}\vec{\beta}})/\epsilon^2$ we get the bound

$$\Pr(|\hat{\chi}_{\vec{\alpha}\vec{\beta}}^{\text{MoM}} - \chi_{\vec{\alpha}\vec{\beta}}| > \epsilon) \leq 2e^{-\frac{K}{2}}. \quad (\text{C14})$$

In order to have an error smaller than ϵ with probability smaller than $1 - \delta$ we need $K = 2 \log(\frac{2}{\delta})$. Putting everything together we get

$$M \sim \log\left(\frac{1}{\delta}\right) \frac{\text{Var}(G_{\vec{\alpha}\vec{\beta}})}{\epsilon^2}. \quad (\text{C15})$$

Therefore we can now state

Lemma C1. Let $\vec{\alpha}, \vec{\beta} \in \{0, 1, 2, 3\}^n$. Then $\chi_{\vec{\alpha}\vec{\beta}}$ can be estimated with error ϵ and probability $1 - \delta$ using only single-qubit state preparation and measurement with

$$\mathcal{O}\left(\frac{\text{Var}(G_{\vec{\alpha}\vec{\beta}})}{\epsilon^2} \log \frac{1}{\delta}\right) = \mathcal{O}\left(\frac{C^d}{\epsilon^2} \log \frac{1}{\delta}\right) \quad (\text{C16})$$

queries to $\mathcal{E}$.

Proof. Putting together Eqs. (C4) and (C15) we obtained the desired result. ■

We now wish to estimate the corresponding channel up to error ϵ in the ℓ_2 -norm. To this end we adapt the proof of Theorem 1 from Ref. [29] to our measurement protocol. A central technical ingredient in that work is a Bohnenblust–Hille type inequality for maps measured in the $S_1 \rightarrow$

S_∞ operator norm; we recall the relevant definitions here for the reader's convenience.

For a matrix M we denote by $\|M\|_{S_1}$ its Schatten—1 (trace) norm and by $\|M\|_{S_\infty}$ its Schatten— ∞ (operator) norm. For a linear superoperator $\Phi : \mathbb{C}^{d \times d} \rightarrow \mathbb{C}^{d \times d}$ we define the induced norm from S_1 to S_∞ as

$$\|\Phi\|_{S_1 \rightarrow S_\infty} := \sup_{M \neq 0} \frac{\|\Phi(M)\|_{S_\infty}}{\|M\|_{S_1}}. \quad (\text{C17})$$

We will also use the induced $S_1 \rightarrow S_1$ norm $\|\Phi\|_{S_1 \rightarrow S_1} := \sup_{M \neq 0} \|\Phi(M)\|_{S_1} / \|M\|_{S_1}$, and the standard fact that for any matrix A one has $\|A\|_{S_\infty} \leq \|A\|_{S_1}$.

For convenience we also report the main technical statement from Ref. [29], which we generalize to trace nonincreasing maps.

Lemma C2 (Bohnenblust–Hille inequality for trace nonincreasing maps). Let $\mathcal{E}$ be a completely positive, trace nonincreasing map of degree at most d , with associated process matrix χ . Then there exists a constant K such that

$$\|\chi\|_{\frac{2d}{d+1}} \leq K^d. \quad (\text{C18})$$

Proof. By Ref. [29] Theorem 12 we have the general bound $\|\chi\|_{2d/(d+1)} \leq K^d \|\mathcal{E}\|_{S_1 \rightarrow S_\infty}$. We are only left to prove that $\|\mathcal{E}\|_{S_1 \rightarrow S_\infty} \leq 1$ for any completely positive, trace nonincreasing map $\mathcal{E}$. Now note that for any such map $\mathcal{E}$ and any operator M ,

$$\|\mathcal{E}(M)\|_1 \leq \|M\|_1. \quad (\text{C19})$$

This is a standard contractivity property of trace nonincreasing CP maps. Therefore, $\|\mathcal{E}\|_{S_1 \rightarrow S_1} \leq 1$. Finally, since

$$\|\mathcal{E}\|_{S_1 \rightarrow S_\infty} \leq \|\mathcal{E}\|_{S_1 \rightarrow S_1}, \quad (\text{C20})$$

we obtain $\|\mathcal{E}\|_{S_1 \rightarrow S_\infty} \leq 1$. Plugging this back into the Bohnenblust–Hille inequality gives the desired bound. ■

We are now ready to prove the theorem that provides the scaling of our tomography protocol.

Theorem C1. Let $\mathcal{E}$ be an n -qubit degree- d quantum channel. There is an algorithm that learns $\mathcal{E}$ in ℓ_2 -distance with precision ϵ and failure probability δ , using only single-qubit state preparation and measurements with

$$e^{\mathcal{O}(d \log(1/\epsilon))} \cdot \mathcal{O}\left(\log\left(\frac{n}{\delta}\right)\right) \quad (\text{C21})$$

queries to $\mathcal{E}$.

Proof. We first state the algorithm. Let K be the constant in the statement of Lemma C2 and C that of Lemma C1.

ALGORITHM 1. Learning low-degree channels with single-qubit state preparation and measurements, via BH inequality.

-
- 1: **Input:** A quantum channel $\mathcal{E}$ of degree at most d , error ϵ , and failure probability δ
 - 2: Let $c = \epsilon^{4d+2} K^{-4d^2}$
 - 3: Perform $M_1 = \mathcal{O}\left(\frac{C^d}{c^2} \log \frac{(4n)^d}{\delta}\right)$ randomized Pauli measurements of $\mathcal{E}$ to estimate $(\chi_{\vec{\alpha}, \vec{\alpha}})_{\vec{\alpha}}$
 - 4: Let $(\chi'_{\vec{\alpha}, \vec{\alpha}})_{\vec{\alpha}}$ be the associated empirical distribution
 - 5: Perform $M_2 = \mathcal{O}\left(\frac{C^d}{c^2 \epsilon^2} \log \frac{1}{c^2 \delta}\right)$ randomized Pauli measurements of $\mathcal{E}$
 - 6: **for** $\vec{\alpha}, \vec{\beta} \in X_c = \{\vec{\alpha} : |\chi'_{\vec{\alpha}, \vec{\alpha}}| \geq c\}$ **do**
 - 7: Approximate $\chi_{\vec{\alpha}, \vec{\beta}}$ with $\chi''_{\vec{\alpha}, \vec{\beta}}$ using Lemma 1
 - 8: **end for**
 - 9: **Output:** $\sum_{\vec{\alpha}, \vec{\beta} \in X_c} \chi''_{\vec{\alpha}, \vec{\beta}} P_{\vec{\alpha}}(\cdot) P_{\vec{\beta}}$
-

Let $c > 0$ be determined later. In the first part of the algorithm we perform $M_1 = \mathcal{O}(C^d/c^2 \log((4n)^d/\delta))$ randomized single-qubit Pauli measurements on $\mathcal{E}$. We use the results to estimate the diagonal of the corresponding process matrix χ . Let $(\chi'_{\vec{\alpha}, \vec{\alpha}})_{\vec{\alpha} \in \{0,1,2,3\}^n}$ be the empirical distribution after M_1 samples. Consider the event,

$$E_1 = \{|\chi_{\vec{\alpha}, \vec{\alpha}} - \chi'_{\vec{\alpha}, \vec{\alpha}}| \leq c, \forall \vec{\alpha} \in \{0, 1, 2, 3\}^n\}. \quad (\text{C22})$$

Then the complementary event is equivalent to

$$\bar{E}_1 = \bigcup_{\vec{\alpha}} \{|\chi_{\vec{\alpha}, \vec{\alpha}} - \chi'_{\vec{\alpha}, \vec{\alpha}}| > c\} := \bigcup_{\vec{\alpha}} \bar{E}_1^{\vec{\alpha}} \quad (\text{C23})$$

So for the union bound

$$\Pr[\bar{E}_1] \leq \sum_{\vec{\alpha}} \Pr[\bar{E}_1^{\vec{\alpha}}] \leq \delta, \quad (\text{C24})$$

where we have used the fact that by Lemma C1, taking $M_1 = \mathcal{O}(C^d/c^2 \log((4n)^d/\delta))$ ensures that $\Pr[\bar{E}_1^{\vec{\alpha}}] < \delta/(4n^d)$ for every $\vec{\alpha} \in \{0, 1, 2, 3\}^n$. Therefore we get that by taking M_1 samples we obtain $\Pr[E_1] \geq 1 - \delta$.

Define $X_c = \{\vec{\alpha} : |\chi'_{\vec{\alpha}, \vec{\alpha}}| \geq c\}$. Since $\sum_{\vec{\alpha} \in X_c} \chi_{\vec{\alpha}, \vec{\alpha}} \leq 1$, we know

$$|X_c| \leq c^{-1}. \quad (\text{C25})$$

Assuming E_1 holds, for $\vec{\alpha} \notin X_c$ we have

$$|\chi_{\vec{\alpha}, \vec{\alpha}}| \leq |\chi'_{\vec{\alpha}, \vec{\alpha}}| + ||\chi_{\vec{\alpha}, \vec{\alpha}} - \chi'_{\vec{\alpha}, \vec{\alpha}}|| \leq 2c. \quad (\text{C26})$$

In particular,

$$\vec{\alpha} \notin X_c \Rightarrow |\chi_{\vec{\alpha}, \vec{\beta}}| \leq \sqrt{|\chi_{\vec{\alpha}, \vec{\alpha}}| |\chi_{\vec{\beta}, \vec{\beta}}|} \leq \sqrt{2c}, \quad \forall \vec{\beta}, \quad (\text{C27})$$

where the first inequality uses that χ is positive semidefinite and the second uses (C26) and $|\chi_{\vec{\beta}, \vec{\beta}}| \leq 1$.

In the second part of the algorithm, by Lemma C1 and with

$$M_2 = \mathcal{O}\left(\frac{C^d}{c^2 \epsilon^2} \log \frac{1}{c^2 \delta}\right), \quad (\text{C28})$$

we can find approximations $\chi''_{\vec{\alpha}, \vec{\beta}}$ of $\chi_{\vec{\alpha}, \vec{\beta}}$ for every $\vec{\alpha}, \vec{\beta} \in X_c$ such that

$$\sup_{(\vec{\alpha}, \vec{\beta}) \in X_c \times X_c} |\chi_{\vec{\alpha}, \vec{\beta}} - \chi''_{\vec{\alpha}, \vec{\beta}}| \leq c\epsilon, \quad (\text{C29})$$

with probability at least $1 - \delta$. In fact let us consider the following event:

$$E_2 = \Pr\left[|\chi_{\vec{\alpha}, \vec{\beta}} - \chi''_{\vec{\alpha}, \vec{\beta}}| \leq c\epsilon, \quad \forall (\vec{\alpha}, \vec{\beta}) \in X_c \times X_c\right]. \quad (\text{C30})$$

Then its complement reads

$$\begin{aligned} & \Pr\left[\bigcup_{(\vec{\alpha}, \vec{\beta}) \in X_c \times X_c} |\chi_{\vec{\alpha}, \vec{\beta}} - \chi''_{\vec{\alpha}, \vec{\beta}}| > c\epsilon\right] \\ & \leq \sum_{(\vec{\alpha}, \vec{\beta}) \in X_c \times X_c} \Pr\left[|\chi_{\vec{\alpha}, \vec{\beta}} - \chi''_{\vec{\alpha}, \vec{\beta}}| > c\epsilon\right]. \end{aligned} \quad (\text{C31})$$

Then by using $M_2 = \mathcal{O}(K^d/c^2 \epsilon^2 \log(1/c^2 \delta))$ from Lemma C1 we get that

$$\Pr[\bar{E}_2] \leq \frac{1}{c^2} c^2 \delta = \delta, \quad (\text{C32})$$

where we have used (C25). Let $\mathcal{E}''(\cdot) = \sum_{\vec{\alpha}, \vec{\beta} \in X_c} \chi''_{\vec{\alpha}, \vec{\beta}} \sigma_{\vec{\alpha}}(\cdot) \sigma_{\vec{\beta}}$. Then,

$$\|\mathcal{E} - \mathcal{E}''\|_2^2 = \sum_{\vec{\alpha}, \vec{\beta} \in X_c} |\chi_{\vec{\alpha}, \vec{\beta}} - \chi''_{\vec{\alpha}, \vec{\beta}}|^2 + \sum_{\vec{\alpha} \vee \vec{\beta} \notin X_c} |\chi_{\vec{\alpha}, \vec{\beta}}|^2 \quad (\text{C33})$$

$$\leq \epsilon^2 + \sum_{\vec{\alpha} \vee \vec{\beta} \notin X_c} \left|\chi_{\vec{\alpha}, \vec{\beta}}\right|^{\frac{2}{d+1}} \left|\chi_{\vec{\alpha}, \vec{\beta}}\right|^{\frac{2d}{d+1}} \quad (\text{C34})$$

$$\leq \epsilon^2 + (2c)^{\frac{1}{d+1}} \|\mathcal{E}\|^{\frac{2d}{d+1}} \quad (\text{C35})$$

$$\leq \epsilon^2 + c^{1/(d+1)} K^d, \quad (\text{C36})$$

where in the equality we have used Parseval's identity; in the first inequality we used Eq. (C25), the learning guarantees of the second part of the algorithm and that $2 = 1/(d+1/2) + 2d/(d+1/2)$; in the second inequality we have used Eq. (C27); and in the third inequality we used the Bohnenblust-Hille inequality for trace nonincreasing maps (Lemma C2).

Hence choosing $c = \epsilon^{2d+2} K^{-d(d+1)}$ and simplifying the expression using $\log(1/\epsilon) \gtrsim d$ yield the desired result. ■

We will now adjust Theorem C1 to reconstruct the channel with recovery guarantee in diamond distance. To do that we will make use of the following result.

Lemma C3. Let $\mathcal{E}$ and $\mathcal{E}'$ be two quantum channels, with process matrix respectively χ and χ' , then the following inequality holds

$$\|\mathcal{E} - \mathcal{E}'\|_{\diamond} \leq \sum_{\vec{\alpha}, \vec{\beta}} |\chi_{\vec{\alpha}\vec{\beta}} - \chi'_{\vec{\alpha}\vec{\beta}}|. \quad (\text{C37})$$

Proof. Let us write the definition of diamond distance between two channels,

$$\|\mathcal{E} - \mathcal{E}'\|_{\diamond} = \sup_{\rho} \|(\mathcal{E} \otimes \mathbb{1})\rho - (\mathcal{E}' \otimes \mathbb{1})\rho\|_1, \quad (\text{C38})$$

where the maximization is done over all density matrices ρ of dimension 2^{2n} . Then the following inequalities hold

$$\begin{aligned} & \|\mathcal{E} - \mathcal{E}'\|_{\diamond} \\ &= \sup_{\rho} \left\| \sum_{\vec{\alpha}, \vec{\beta}} (\chi_{\vec{\alpha}\vec{\beta}} - \chi'_{\vec{\alpha}\vec{\beta}}) (P_{\vec{\alpha}} \otimes \mathbb{1}) \rho (P_{\vec{\beta}} \otimes \mathbb{1}) \right\|_1 \end{aligned} \quad (\text{C39})$$

$$\leq \sum_{\vec{\alpha}, \vec{\beta}} |\chi_{\vec{\alpha}\vec{\beta}} - \chi'_{\vec{\alpha}\vec{\beta}}| \sup_{\rho} \|(P_{\vec{\alpha}} \otimes \mathbb{1}) \rho (P_{\vec{\beta}} \otimes \mathbb{1})\|_1 \quad (\text{C40})$$

$$\leq \sum_{\vec{\alpha}, \vec{\beta}} |\chi_{\vec{\alpha}\vec{\beta}} - \chi'_{\vec{\alpha}\vec{\beta}}| \sup_{\rho} \|(P_{\vec{\alpha}} \otimes \mathbb{1})\|_{\infty} \|\rho\|_1 \|(P_{\vec{\beta}} \otimes \mathbb{1})\|_{\infty}, \quad (\text{C41})$$

where from line one to two we have used the triangular inequality and from line two to three we have used Hölder's inequality. Lastly, knowing that the infinity norm of $P_{\vec{\alpha}} \otimes \mathbb{1}$ and the one norm of ρ are one we can finally write

$$\|\mathcal{E} - \mathcal{E}'\|_{\diamond} \leq \sum_{\vec{\alpha}, \vec{\beta}} |\chi_{\vec{\alpha}\vec{\beta}} - \chi'_{\vec{\alpha}\vec{\beta}}|, \quad (\text{C42})$$

which concludes the proof. ■

We are now ready to state the algorithm for reconstruction in diamond distance.

Theorem C2. Let $\mathcal{E}$ be an n -qubit degree- d quantum channel. There is an algorithm that learns $\mathcal{E}$ in diamond distance with precision ϵ and failure probability δ , using only single-qubit state preparation and measurements with

$$n^{\mathcal{O}(d)} e^{\mathcal{O}(\log(1/\epsilon))} \cdot \mathcal{O}\left(\log\left(\frac{1}{\delta}\right)\right) \quad (\text{C43})$$

queries to $\mathcal{E}$.

Proof. The proof closely follows the one of Theorem C1. The only difference is that now we will need

$$M_2 = \mathcal{O}\left(\frac{C^d}{c^4 \epsilon^2} \log \frac{1}{c^2 \delta}\right). \quad (\text{C44})$$

Then, using Lemma C3 we can estimate the diamond distance as

$$\|\mathcal{E} - \mathcal{E}''\|_{\diamond} \leq \sum_{\vec{\alpha}, \vec{\beta} \in X_c} |\chi_{\vec{\alpha}, \vec{\beta}} - \chi''_{\vec{\alpha}, \vec{\beta}}| + \sum_{\vec{\alpha} \vee \vec{\beta} \notin X_c} |\chi_{\vec{\alpha}, \vec{\beta}}| \quad (\text{C45})$$

$$\leq \epsilon + \sqrt{2c} (4n)^d, \quad (\text{C46})$$

where we have used (C27) to bound the second term along side the fact that a d -degree channel has at most $(4n)^d$ nonzero elements. Therefore by choosing $c = \frac{\epsilon^2}{2} (4n)^{-2d}$ we conclude the proof. ■

1. Unitary acting on polylogarithmic number of qubits

Let us consider a unitary map $\mathcal{U}$ that only acts on a set Q of qubits of cardinality q . Then, given the product structure of the G function, we will have that the rotated $G^{\mathcal{U}}$ will be different only on the qubits on which $\mathcal{U}$ is acting. For this we can write

$$G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s}) = G_{\vec{\gamma}\vec{\delta}}^Q \left(\prod_{i \in \bar{Q}} g_{\gamma_i \delta_i}(r_i, s_i) \right), \quad (\text{C47})$$

where we have called $G_{\vec{\gamma}\vec{\delta}}^Q$ the part of the function with support on Q and $\bar{Q}$ the set of idle qubits. Similarly we know that $\chi_{\vec{\alpha}\vec{\beta}}^c$ will differ from $\chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}}$ on at most q qubits for every $\vec{\alpha}, \vec{\beta}$ so it will have nonzero entries for Pauli strings that have weight at most $d+q$. So, writing down expression

(32) for this case we have

$$\sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^c \left(\sum_{\vec{r}_Q, \vec{s}_Q} \left| G_{\vec{\gamma}\vec{\delta}}^Q(\vec{r}_Q, \vec{s}_Q) \right|^2 \prod_{i \in Q} f_{\alpha_i \beta_i}(r_i, s_i) \right) \times \left(\prod_{i \in \bar{Q}} \sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \right), \quad (\text{C48})$$

where we have defined $\vec{r}_Q, \vec{s}_Q$ as the part of the input and output state that has support on Q . If we call K the value of the first parenthesis, we can write from Eq. (30),

$$\text{Var}(G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s})) \leq K \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^c \left(\prod_{i \in \bar{Q}} \sum_{r_i, s_i} |g_{\gamma_i \delta_i}(r_i, s_i)|^2 f_{\alpha_i \beta_i}(r_i, s_i) \right) \quad (\text{C49})$$

$$\leq KC^{4d} \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^c \delta_{\vec{\alpha}\vec{\beta}, \vec{0}} \quad (\text{C50})$$

$$\leq KC^{4d} \left(\sum_{\vec{\alpha}\vec{\beta}, \vec{0}} \chi_{\vec{\alpha}\vec{\beta}, \vec{0}}^c + 1 \right) \quad (\text{C51})$$

$$\leq (4^q + 1)KC^{4d} \leq (K')^q C^{4d}, \quad (\text{C52})$$

where from line one to two we have used once again that the number of qubits on which $\alpha_i, \beta_i, \gamma_i$ or δ_i differ from the identity is at most $4d$ on $\bar{Q}$ and we have used condition (ii) defined in Appendix C. From line two to three we have used the fact that the trace of χ is smaller equal than 1 and from line three to four we have used the fact that the process matrix is positive semidefinite, so each off-diagonal term is bounded by the average of the corresponding diagonal terms ($\chi_{\vec{\alpha}\vec{\beta}} \leq (\chi_{\vec{\alpha}\vec{\alpha}} + \chi_{\vec{\beta}\vec{\beta}})/2$) so

$$\sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}} \leq \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\alpha}} \leq 4^q. \quad (\text{C53})$$

At last we defined K' as $5^q K^{1/q}$ and got the final inequality. We can now see that if we take $q = \mathcal{O}(\text{poly}(\log(n)))$ we would get that the scaling of the variance becomes now polynomial in n , instead of constant

$$\text{Var}(G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s})) = \mathcal{O}(e^q e^d) \rightarrow \mathcal{O}(n^p e^d), \quad (\text{C54})$$

where p is the product of the exponents appearing in the polynomial dependence on log of the support of such gates. If we accept this trade-off we would be capable of performing efficient process tomography on any gate acting on a polylogarithmic number of qubits, via Eq. (C15).

2. One layer of two-qubit gates

As mentioned above we are interested in benchmarking the two-qubit gates of our device, so $\mathcal{U}$ will be a layer of two-qubit gates. We will therefore define $n/2$ unitaries $\mathcal{U}_{\mathbf{j}}$ each one acting on a pair of qubits $\mathbf{j} = (j_0, j_1)$. We will indicate with $\vec{r}_{\mathbf{j}}$ and $\vec{s}_{\mathbf{j}}$ the qubits on which $\mathcal{U}_{\mathbf{j}}$ is acting. We will call J the set of pair of qubits such that no qubit is ever found in more than one pair at a time. This set has of course cardinality $n/2$.

As mentioned in the main text $G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s})$ can now be written as a product of $n/2$ two-qubit functions,

$$G_{\vec{\alpha}\vec{\beta}}^{\mathcal{U}}(\vec{r}, \vec{s}) = \prod_{\mathbf{j} \in J} g_{\mathcal{U}_{\mathbf{j}}}^{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}), \quad (\text{C55})$$

where we have defined the two qubit rotated $\mathbf{g}$ function as $g_{\mathcal{U}_{\mathbf{j}}}^{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}})$. Let us now dive into more details of the derivations of the minimization procedure. In this setting we can write expression (32) as

$$\frac{1}{18^n} \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^{\varepsilon} \prod_{\mathbf{j} \in J} \left(\sum_{\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}} |g_{\mathcal{U}_{\mathbf{j}}}^{\vec{\gamma}_{\mathbf{j}} \vec{\delta}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}})|^2 \langle \vec{r}_{\mathbf{j}} | \sigma_{\alpha_{\mathbf{j}}} \mathcal{U}_{\mathbf{j}}(|\vec{s}_{\mathbf{j}}\rangle \langle \vec{s}_{\mathbf{j}}|) \sigma_{\beta_{\mathbf{j}}} | \vec{r}_{\mathbf{j}} \rangle \right). \quad (\text{C56})$$

Since $\chi_{\vec{\alpha}\vec{\beta}}^{\varepsilon}$ is nonzero only of for the Pauli operators with degree at most d , to have variance not scaling exponentially with system size, we would like the term in bracket to be bounded by 1 for $\alpha_{\mathbf{j}}, \beta_{\mathbf{j}}, \gamma_{\mathbf{j}}, \delta_{\mathbf{j}} = 0$,

$$\sum_{\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}} |g_{00}^{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}})|^2 \langle \vec{r}_{\mathbf{j}} | \mathcal{U}_{\mathbf{j}}(|\vec{s}_{\mathbf{j}}\rangle \langle \vec{s}_{\mathbf{j}}|) | \vec{r}_{\mathbf{j}} \rangle = \sum_{\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}} |g_{00}^{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}})|^2 p_{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}) \quad (\text{C57})$$

$$= \sum_{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}} \chi_{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}^{\mathcal{U}_{\mathbf{j}}} \sum_{\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}} |g_{00}^{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}})|^2 f_{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}), \quad (\text{C58})$$

where we have defined $p_{\mathcal{U}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}})$ as $(1/18^2) \langle \vec{r}_{\mathbf{j}} | \mathcal{U}_{\mathbf{j}}(|\vec{s}_{\mathbf{j}}\rangle \langle \vec{s}_{\mathbf{j}}|) | \vec{r}_{\mathbf{j}} \rangle$. Now $f_{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}(\vec{r}_{\mathbf{j}}, \vec{s}_{\mathbf{j}}) := f_{\alpha_{j_0} \beta_{j_0}}(r_{j_0}, s_{j_0}) \otimes f_{\alpha_{j_1} \beta_{j_1}}(r_{j_1}, s_{j_1})$ acts on two qubits and can be thought as a (1296×256) matrix, similarly to what done before. It is now possible to define the left kernel of $f_{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}$ as a (1040×1296) matrix $\mathbf{K}$ and again we will have that $g_{\mathcal{U}_{\mathbf{j}}}^{\vec{\gamma}_{\mathbf{j}} \vec{\delta}_{\mathbf{j}}} + \vec{x}^T \mathbf{K}$ is the inverse of $f_{\vec{\alpha}_{\mathbf{j}} \vec{\beta}_{\mathbf{j}}}$. We are now left with minimizing (C57) for the 1040 free

parameters encoded in $\vec{x}$,

$$\min_{\vec{x}} \sum_{\vec{r}_j, \vec{s}_j} |g_{\mathcal{U}_j}^{00} + \vec{x}^T K|^2 p_{\mathcal{U}_j}(\vec{r}_j, \vec{s}_j). \quad (\text{C59})$$

This again is a quadratic minimization problem with a well-known analytic solution. We find that this term is bounded by 1 if $\mathcal{U}$ is a Clifford gate,

$\mathcal{U} \in \text{Clifford}$

$$\implies \min_{\vec{x}} \sum_{\vec{r}_j, \vec{s}_j} |g_{\mathcal{U}_j}^{00}(\vec{r}_j, \vec{s}_j) + \vec{x}^T K|^2 p_{\mathcal{U}_j}(\vec{r}_j, \vec{s}_j) = 1. \quad (\text{C60})$$

The code to get the minimized $g_{\mathcal{U}_j}$ for a generic two-qubit unitary is publicly available in Ref. [37]. In the same location it is possible to find the explicit solution for the ISWAP gate as well as the CNOT. Given this result let us now give an analytic proof of sample complexity.

3. Proof of sample complexity for a layer of two-qubit gates

When the bound in Eq. (C60) holds, we can bound the variance. Let us call C the constant such that $\sum_{\vec{r}_j, \vec{s}_j} |g_{\mathcal{U}_j}^{\gamma_j \delta_j}(\vec{r}_j, \vec{s}_j)|^2 \langle \vec{r}_j | \sigma_{\alpha_j} \mathcal{U}_j(|\vec{s}_j\rangle \langle \vec{s}_j|) \sigma_{\beta_j} | \vec{r}_j \rangle \leq C$ for every $\alpha_j, \beta_j, \gamma_j, \delta_j$. Then we rewrite Eq. (30) as

$$\text{Var}(G_{\vec{\gamma}\vec{\delta}}) \leq \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}} \sum_{\vec{r}, \vec{s}} |G_{\vec{\gamma}\vec{\delta}}^{\mathcal{U}}(\vec{r}, \vec{s})|^2 \langle \vec{r} | \sigma_{\vec{\alpha}} \mathcal{U}(|\vec{s}\rangle \langle \vec{s}|) \sigma_{\vec{\beta}} | \vec{r} \rangle \quad (\text{C61})$$

$$= \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}} \prod_{j \in J} \left(\sum_{\vec{r}_j, \vec{s}_j} |g_{\mathcal{U}_j}^{\gamma_j \delta_j}(\vec{r}_j, \vec{s}_j)|^2 \langle \vec{r}_j | \sigma_{\alpha_j} \mathcal{U}_j(|\vec{s}_j\rangle \langle \vec{s}_j|) \sigma_{\beta_j} | \vec{r}_j \rangle \right) \quad (\text{C62})$$

$$\leq C^{4d} \sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}}, \quad (\text{C63})$$

where from line two to three we have used the fact that the indexes $\alpha_j, \beta_j, \gamma_j, \delta_j$ differ from the identity on at most $4d$ qubits and for the rest the term in bracket is equal to 1. We are now left to bound the sum of the entries of the process matrix χ . To do that we will make use of Lemma C2 and of the known properties of p -norms: $\|x\|_r \leq n^{1/r-1/p} \|x\|_p$ for $1 \leq r < p$. By choorsing $r = 1$ and $p = 2d/(d+1)$ we get the following bound:

$$\sum_{\vec{\alpha}\vec{\beta}} \chi_{\vec{\alpha}\vec{\beta}}^{\mathcal{E}} \leq \|\chi^{\mathcal{E}}\|_1 \leq n^{1-\frac{d+1}{2d}} \|\chi^{\mathcal{E}}\|_{\frac{2d}{d+1}} \leq nK^d. \quad (\text{C64})$$

Therefore we obtain

$$\text{Var}(G_{\vec{\gamma}\vec{\delta}}) \leq nC^{4d}K^d. \quad (\text{C65})$$

Making use of Eq. (C15) we are able to do efficient post-processing of a layer of two-qubit Clifford gates with just a linear overhead.

If now we want to estimate the variance of the function $G_{\mathcal{U}}$ corresponding to a layer of two-qubit Clifford gates and some non-Clifford gates, we need to put together the content of this section with Appendix C 1. Given the structure of the G matrix we can split the function in its Clifford and non-Clifford component,

$$G_{\mathcal{U}} = G_{\mathcal{U}_{\text{Cl}}} \cdot G_{\mathcal{U}_{\text{non-Cl}}}. \quad (\text{C66})$$

Then we can write

$$\text{Var}(G_{\mathcal{U}}) \leq \text{Var}(G_{\mathcal{U}_{\text{Cl}}}) \cdot \text{Var}(G_{\mathcal{U}_{\text{non-Cl}}}) \leq ne^q C^d, \quad (\text{C67})$$

where q is the number of qubits on which the non-Clifford gates are acting. Therefore as long as the number of qubits on which the non-Clifford gates are acting is polylogarithmic we have a variance of a single entry scaling polynomially in system size.

If we want to estimate the reconstructed channel in ℓ_2 -norm and diamond norm we need to adapt the proof of Theorem C1 to take into account the different scaling of the variance. This brings as to state the following Lemma.

Lemma C4. Let $\mathcal{E}$ be an n -qubit degree- d quantum channel, preceded by a gate layer composed of arbitrary unitaries with total support on at most q qubits, and two-qubit Clifford gates acting on the remaining qubits. Let $\mathcal{C}$ be the composition of the gate layer and $\mathcal{E}$. There is an algorithm that learns $\mathcal{E}$ in ℓ_2 -distance with precision ϵ and failure probability δ , using only single-qubit state preparation and measurements with

$$e^{\mathcal{O}(d \log(1/\epsilon))} \cdot \mathcal{O}\left(ne^q \log\left(\frac{n}{\delta}\right)\right) \quad (\text{C68})$$

queries to $\mathcal{C}$. Similarly there is an algorithm that learns $\mathcal{E}$ in diamond distance with precision ϵ and failure probability δ , using only single-qubit state preparation and measurements with

$$n^{\mathcal{O}(d)} e^{\mathcal{O}(\log(1/\epsilon)) + q} \cdot \mathcal{O}\left(\log\left(\frac{1}{\delta}\right)\right). \quad (\text{C69})$$

Proof. The proof is a straightforward adaptation from Theorem C1 and it only requires substituting C^d with $ne^q C^d$. Now we would have $M_1 = \mathcal{O}((ne^q C^d/c^2) \log((4n)^d/\delta))$ and $M_2 = \mathcal{O}((ne^q C^d/c^2 \epsilon^2) \log(1/c^2 \delta))$. This translates to a final sample

complexity of

$$M = e^{\mathcal{O}(d \log(1/\epsilon))} \cdot \mathcal{O}\left(ne^q \log\left(\frac{n}{\delta}\right)\right). \quad (\text{C70})$$

Doing the same substitution in Theorem C2 we get the desired scaling,

$$M = n^{\mathcal{O}(d)} e^{\mathcal{O}(\log(1/\epsilon)) + q} \cdot \mathcal{O}\left(\log\left(\frac{1}{\delta}\right)\right). \quad (\text{C71})$$

Therefore we can see that restricting to a polylogarithmic number of non-Clifford gates, we would get $e^q \rightarrow n^p$ where p is related to the degree of the polynomial in log number of non-Clifford gates. This preserves the efficiency of the protocol while introducing a polynomial factor in system size. ■

4. Numerical analysis

Since for the case of sampling directly of the error channel $\mathcal{E}$ the scaling is constant for every entry of χ we suspect that the linear bound that we get in this case is not tight. To verify that, we consider a layer of two-qubit gates followed by the single-qubit noise model defined in Eq. (47). We then compute the variance of the function $G^{\mathcal{U}}$ that in this case is a product of two-qubit functions $\mathbf{g}_{\text{ISWAP}}$. In Fig. 5 is reported the numerical computation of the variance of the function $G_{00}^{\mathcal{U}}$ (that is the entry corresponding to all identities) for increasing system size. We can see that the value saturates to a constant, similar to what would happen without classical postprocessing. So, it is possible that even though the best bound we could find is linear, the scaling is actually constant in some relevant cases.

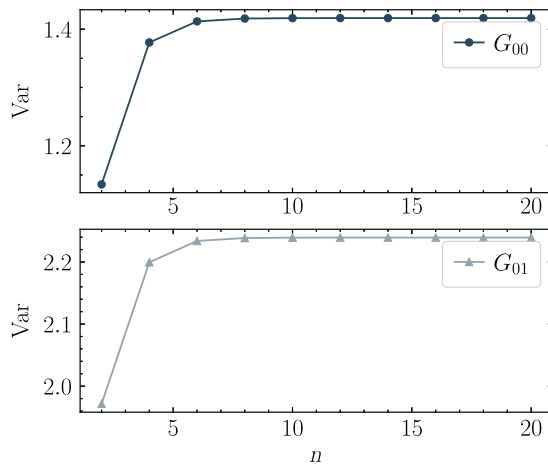


FIG. 5. Variance of the entry of $\mathbf{G}$ when a layer of ISWAP gates is applied, followed by an error channel with error parameters decaying exponentially as defined in Eq. (47). The computation is carried for the entries corresponding to $\vec{\alpha} = \vec{\beta} = (0, \dots, 0)$ (top panel) and $\vec{\alpha} = (0, \dots, 0)$, $\vec{\beta} = (x, 0, \dots, 0)$ (bottom panel).

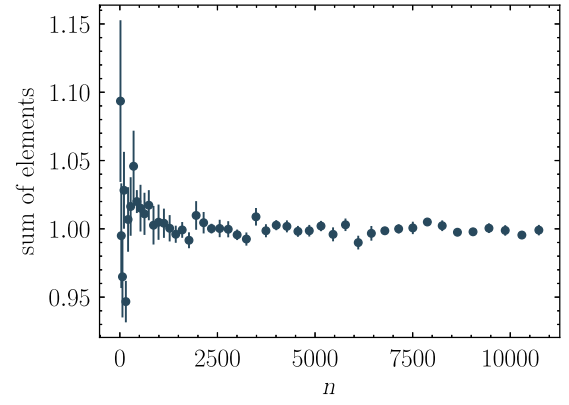


FIG. 6. Sum of elements of random positive semidefinite matrices for increasing system size. The average is taken over 10 independent samples and the error bars are computed as the standard deviation.

Given the previous results, we investigated under which conditions the chain of inequalities ceases to be tight. Specifically, we analyzed the behavior of the bounds going from Eqs. (C63) to (C65). To this end, we generated random positive semidefinite matrices of increasing dimension. Specifically for every n we constructed matrices of dimension $\sim (n^2 \times n^2)$ to simulate the scaling with system size of a two-degree Pauli channel. We observed that the sum of their elements does not grow with the system size, as shown by the results summarized in Fig. 6. While we are aware that this behavior does not hold for all positive semidefinite matrices, our findings suggest that, in many practically relevant scenarios, the scaling may indeed remain constant.

Given the encouraging results of the previous section, where numerical computations exhibited better scaling with respect to analytical bounds, we computed the variance of $G^{\min}$ for a layer of non-Clifford gates, while

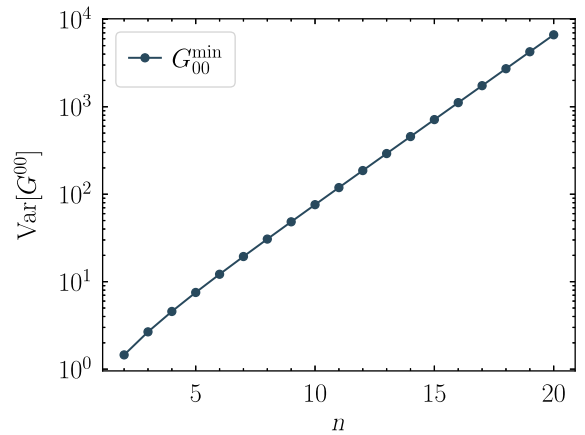


FIG. 7. Variance of the entry G_{00} when a layer of T gates is applied, followed by an error channel with error parameters decaying exponentially defined in Eq. (47).

maintaining the same error model as in Sec. V. More specifically we used a layer of T gates,

$$T = \begin{pmatrix} 1 & 0 \\ 0 & \exp\left(\frac{i\pi}{4}\right) \end{pmatrix}. \quad (\text{C72})$$

Results are reported in Fig. 7 and show exponential scaling.

-
- [1] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, *Nature* **618**, 500 (2023).
 - [2] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Japan* **90**, 032001 (2021).
 - [3] A. Seif, Z.-P. Ciani, S. Zhou, S. Chen, and L. Jiang, Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors, *PRX Quantum* **4**, 010303 (2023).
 - [4] B. M. Terhal, Quantum error correction for quantum memories, *Rev. Mod. Phys.* **87**, 307 (2015).
 - [5] 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).
 - [6] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
 - [7] J. F. Poyatos, J. I. Cirac, and P. Zoller, Complete characterization of a quantum process: The two-bit quantum gate, *Phys. Rev. Lett.* **78**, 390 (1997).
 - [8] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan, S. M. Clark, R. Blume-Kohout, I. Siddiqi, and K. C. Young, Experimental characterization of crosstalk errors with simultaneous gate set tomography, *PRX Quantum* **2**, 040338 (2021).
 - [9] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
 - [10] 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).
 - [11] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, *J. Mod. Opt.* **44**, 2455 (1997).
 - [12] T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta, Projected least-squares quantum process tomography, *Quantum* **6**, 844 (2022).
 - [13] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of addressability by simultaneous randomized benchmarking, *Phys. Rev. Lett.* **109**, 240504 (2012).
 - [14] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, *Phys. Rev. Res.* **6**, 013029 (2024).
 - [15] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, Correlated randomized benchmarking, *arXiv:2003.02354*.
 - [16] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, General framework for randomized benchmarking, *PRX Quantum* **3**, 020357 (2022).
 - [17] H.-Y. Huang, S. Chen, and J. Preskill, Learning to predict arbitrary quantum processes, *PRX Quantum* **4**, 040337 (2023).
 - [18] Y. Wang, G. Liu, Z. Liu, Y. Tang, X. Ma, and H. Dai, Robust estimation of nonlinear properties of quantum processes, *Phys. Rev. A* **110**, 032415 (2024).
 - [19] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. H. Werner, J. Eisert, and I. Roth, Shadow estimation of gate-set properties from random sequences, *Nat. Commun.* **14**, 5039 (2023).
 - [20] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Shadow process tomography of quantum channels, *Phys. Rev. A* **107**, 042403 (2023).
 - [21] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Detecting crosstalk errors in quantum information processors, *Quantum* **4**, 321 (2020).
 - [22] N. Kaufmann, I. Rojko, and F. Reiter, Characterization of coherent errors in gate layers with robustness to Pauli noise, *Phys. Rev. Appl.* **23**, 034014 (2025).
 - [23] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, *Nat. Phys.* **16**, 1184 (2020).
 - [24] C. Rouzé and D. Stilck Franca, Efficient learning of the structure and parameters of local Pauli noise channels, *arXiv:2307.02959* [quant-ph].
 - [25] S. T. Flammia and R. O'Donnell, Pauli error estimation via population recovery, *Quantum* **5**, 549 (2021).
 - [26] L. C. G. Govia, G. J. Ribeill, D. Ristè, M. Ware, and H. Krovi, Bootstrapping quantum process tomography via a perturbative ansatz, *Nat. Commun.* **11**, 1084 (2020).
 - [27] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, Quantum process tomography with unsupervised learning and tensor networks, *Nat. Commun.* **14**, 2858 (2023).
 - [28] C. Wadhwa, L. Lewis, E. Kashefi, and M. Doosti, Agnostic process tomography, *arXiv:2410.11957*.
 - [29] S. Arunachalam, A. Dutt, F. E. Gutiérrez, and C. Palazuelos, Learning low-degree quantum objects, *arXiv:2405.10933* [quant-ph].
 - [30] Notice that there is a degree of arbitrariness in defining $\mathcal{E}_{\mathcal{U}} := \mathcal{C}_{\mathcal{U}} \circ \mathcal{U}^{\dagger}$ or $\mathcal{E}'_{\mathcal{U}} := \mathcal{U}^{\dagger} \circ \mathcal{C}_{\mathcal{U}}$. The two representations are connected via $\mathcal{E}_{\mathcal{U}} = \mathcal{U} \circ \mathcal{E}'_{\mathcal{U}} \circ \mathcal{U}^{\dagger}$.
 - [31] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
 - [32] J. Malmi, K. Korhonen, D. Cavalcanti, and G. García-Pérez, Enhanced observable estimation through classical optimization of informationally overcomplete measurement data: Beyond classical shadows, *Phys. Rev. A* **109**, 062412 (2024).
 - [33] A. Caprotti, J. Morris, and B. Dakić, Optimizing quantum tomography via shadow inversion, *Phys. Rev. Res.* **6**, 033301 (2024).

- [34] L. E. Fischer, T. Dao, I. Tavernelli, and F. Tacchino, Dual-frame optimization for informationally complete quantum measurements, *Phys. Rev. A* **109**, 062415 (2024).
- [35] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018 (Association for Computing Machinery, New York, NY, 2018), pp. 325–338.
- [36] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, *Phys. Rev. A* **97**, 032306 (2018).
- [37] M. Crupi, noise_characterization, 2025, https://github.com/m-crupi/noise_characterization.git
- [38] H. Yu and T.-C. Wei, Efficient separate quantification of state preparation errors and measurement errors on quantum computers and their mitigation, *Quantum* **9**, 1724 (2025).
- [39] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [40] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [41] H. Jnane, J. Steinberg, Z. Cai, H. C. Nguyen, and B. Koczor, Quantum error mitigated classical shadows, *PRX Quantum* **5**, 010324 (2024).