

Every Benchmark All at Once

Ana Silva and Eliska Greplova¹

¹*QuTech and Kavli Institute of Nanoscience, Delft University of Technology**

As quantum technology matures, the efficient benchmarking of quantum devices remains a key challenge. Although sample-efficient, information-theoretic benchmarking techniques have recently been proposed, there is still a gap in adapting these techniques to contemporary experiments. In this work, we re-formulate five of the most common randomized benchmarking techniques in the modern language of the gate-set shadow tomography. This reformulation brings along several concrete advantages over conventional formulations of randomized benchmarking. For standard and interleaved randomized benchmarking, we can reduce the required gate-set size and, using median-of-means estimators, also reduce the required experimental sample size. For simultaneous and correlated randomized benchmarking, we can additionally reconstruct the Pauli-terms of correlated noise channels using additional post-processing of only a single experimental dataset. We also present a minimal approach to extract leakage errors. Our work provides a clear avenue for comprehensive, reliable, and convenient benchmarking of quantum devices, with all methods formulated under a single umbrella technique that can be easily adapted to a range of experimental quantities and gate sets.

I. INTRODUCTION

The ability to build high performance quantum processors relies on acquiring precise control over quantum systems. To this end, the capacity of performing targeted operations that may only address parts of the system is crucial. However, near-term quantum devices are prone to errors of multi-factorial origin. These may stem, for example, from imperfect manufacturing, inaccuracies in control fields, unforeseen interactions that couple different parts of the system, or undesired interactions with the environment. While quantum error correcting codes have been developed that allow for fault-tolerant quantum computing [1–3], their efficacy requires gate errors to be bounded below a certain threshold [1, 4, 5]. Additionally, error correcting schemes typically assume errors to be uncorrelated [1, 6]. Given these constraints, and the myriad of effects leading to noise, diagnostic tools that provide reliable noise characterization are essential to the development of high performance quantum processors.

Quantum process tomography is a well-established technique that provides full gate error reconstruction [7, 8]. Its implementation, however, is challenging, with two issues arising as limitations: firstly, the measurement overhead required for tomographic reconstruction scales exponentially in the number of qubits [9], and secondly, its gate noise estimations are vulnerable to inaccuracies arising from state preparation and measurement errors (SPAM) [10]. A commonly employed characterization technique that addresses both these difficulties is randomized benchmarking (RB) [11–13]. By forgoing full error reconstruction, RB allows for efficiently probing average noise metrics, while still guaranteeing that these estimates are independent of SPAM errors. To gain information on specific average noise properties, different RB

protocols have been developed that tailor to efficiently extract a targeted noise metric. This can be, for example, learning the noise affecting a specific gate in the gate-set (interleaved RB) [14, 15], estimating the coherence of the noise (purity RB) [16, 17], or estimating crosstalk (simultaneous RB and crosstalk RB) [18, 19].

Beyond error characterization, the dimensionality curse is a shared obstacle to all practical attempts for learning properties of an unknown quantum state. Like with noise reconstruction, full state tomography also scales exponentially in the number of qubits. Hence, its practical implementation has also shifted towards acquiring only partial information on targeted quantum state properties, rather than aiming for full reconstruction. Recently, classical shadow tomography has been introduced as a sample efficient protocol that, albeit not aiming for full state reconstruction, still allows for estimating multiple properties of a quantum state from the same data in post-processing [20]. Crucially, the required number of samples to attain a certain estimation precision is not directly dependent on the number of qubits, but rather on the type of observables we wish to learn [20]. The classical shadow protocol relies on data collected from several repetitions of state evolution through randomly selected unitary gates. Unlike RB, however, the end result in classical shadow tomography is a data-set that allows for the estimation of multiple expectation values, instead of a single figure of merit. Given that knowledge of many noise facets are actually desirable to render a more comprehensive picture of the noise corrupting quantum operations in a real device, it would be desirable to import the multiple observable estimation capabilities of classical shadow to the realm of RB techniques. The gate-set shadow protocol is a recent noise characterization protocol that allows to re-frame RB into a similar framework as the classical shadow protocol [21, 22]. While the protocol establishes an important development in the field of RB techniques, adopting this new approach can be hindered by unclarity on how

* A.C.OliveiraSilva@tudelft.nl

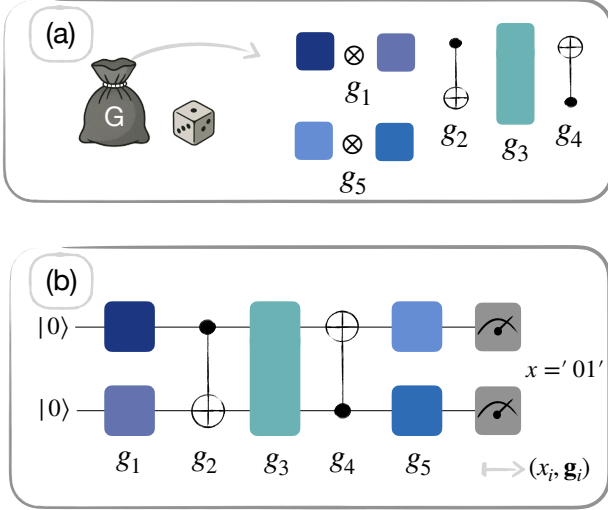


FIG. 1. Overview of the the gate-set shadow protocol: data collection phase. (a) We start by selecting gates uniformly at random from a given gate-set G . (b) This creates a sequence of random gates that are then applied to the input state. The resulting outcome is measured in a computational basis. Iterating over this process allows us to create a data-set formed by the pairing of random sequences with the corresponding outcomes, $(x_i, \mathbf{g}_i)$

it relates and allows extraction of information akin to the already well-established RB protocols.

In this work, we provide several use cases of the gate-set shadow protocol that emulate several existing RB methods. We also explore how the gate-set shadow protocol can be used to provide further complementary noise metrics to those of RB. Specifically, we cover the relation between the gate-set shadow protocol and the RB techniques: standard RB, interleaved RB, simultaneous and correlated RB and leakage RB. Furthermore, we make use of different gate-sets for the implementation of these protocols. Although RB techniques are commonly associated with using Clifford gates, this is not a necessary fixed constraint [15, 23, 24]. In fact, depending on the native gates of a quantum device, opting for a smaller gate-set maybe desirable [25], particularly when implementing arbitrary Clifford gates results in large sequences of native gates. However, uncertainty on how to adapt the protocol for different gate-sets can effectively prevent exploiting this extra degree of freedom. While character RB [26, 27] is a known route to implement RB with general groups as gate-sets, we argue that the shadow gate-set offers a much more unified and practical framework to achieve RB diagnostics with arbitrary gate-sets. We complement our use cases with open source code that can serve as basis for practical implementations of the protocol [28].

II. GATE-SET SHADOW PROTOCOL

The gate set shadow protocol shares many similarities with RB-like methods, but adds to it the flexibility and data driven power of classical shadow tomography, which can simultaneously estimate multiple quantities from the same data. [20]. Similar to RB, we apply a random sequence of m gates. This is followed by a measurement in a computational basis to retrieve a bit-string outcome, x_i . Thus, the fundamental units of information are composed by the pairs $(x_i, \mathbf{g}_i)$, i.e. the bit-string outcome x_i resulting from applying the random sequence of gates $\mathbf{g}_i = \{g_1, g_2, \dots, g_m\}$. A complete dataset is then made of $S = NK$ samples of pairs $(x_i, \mathbf{g}_i)$ where K is the number of randomizations for each sequences length and N is the number of different sequence lengths. Experimentally, the protocol is therefore almost identical to RB, with the difference that the end recovery gate inverting the sequences is removed (see Fig. 1 for an overview of the experimental phase of the protocol).

To describe the measurement process in the post-processing step [20, 21], we employ the language of positive-operator-valued measures (POVMs). POVM measurements are characterized by the POVM elements, $\{E_x\}_x$, each corresponding to a measurement outcome x . These are positive operators whose sum produces the identity, i.e. $\sum_x E_x = \mathbb{1}$. The probability to obtain outcome x is given by $p(x) = \text{Tr}(E_x \rho)$ for a state ρ . In a broad sense, we seek to estimate several expectation values, reflecting crucial properties about our quantum gates. These properties may be the overall error-rate of a gate-set, or more detailed metrics relating to the presence of crosstalk, leakage from a computational subspace, etc. To maintain immunity to SPAM errors, we extract these quantities from an exponential fit to the average sequence functions, similar to RB. In the gate-set shadow protocol, the sequence functions that we are interested in, termed sequence correlation functions [21], are given by

$$f_A(x_i, \mathbf{g}_i) = c_{f_A}^{(0)} \text{Tr}(E_{x_i} S_A \rho S_A^\dagger) \quad (1)$$

where $c_{f_A}^{(0)}$ is a normalization constant, E_{x_i} is the POVM element corresponding to the bit-string x_i , and ρ is the initial state. Additionally, we define the matrix

$$S_A = U_m A U_{m-1} A \dots A U_2 A U_1, \quad (2)$$

where U_i is the ideal noise-free unitary operator corresponding to the gate g_i . Here, A is a user-defined operator (referred to as *probe operator*) and S_A is formed by taking a sequence of gates, inserting the probe operator A between every two consecutive gates, U_i and U_{i+1} . In practice, we evaluate Eq. (1) using the the Pauli transfer matrix (PTM) formalism [29, 30] (see Appendix A 1 for a more detailed description of the PTM formalism). In this formalism, we can write the sequence correlation

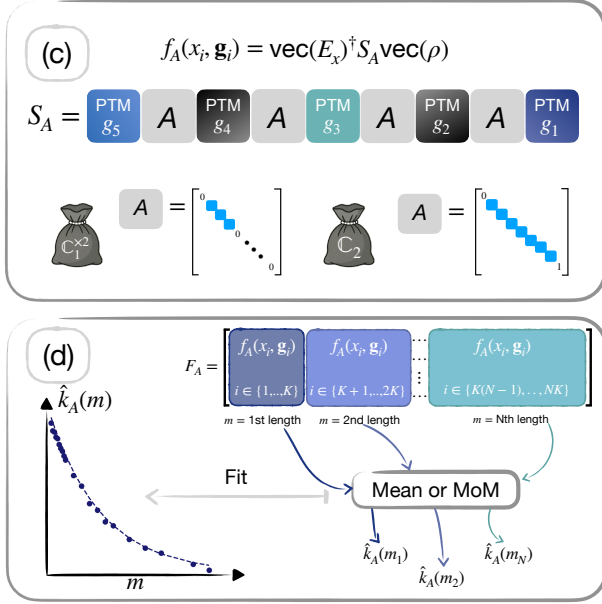


FIG. 2. Overview of the the gate-set shadow protocol: post-processing phase. (c) For each data-set entry $(x_i, \mathbf{g}_i)$, we construct the sequence correlation function f_A . A typical choice for A is to fix it as a subspace projector. Specific choices of subspace projectors are illustrated in the figure for the case of the two-qubit Clifford group (C_2) and the local Clifford group ($C_1^{\times 2}$) as gate-sets. (d) Taking the mean (or median-of-means, MoM) of all the sequence correlation functions gives an estimate of the sequence function, $\hat{k}_A(m)$, which can then be fitted to an exponential decay model to extract targeted fidelities.

function as

$$f_A(x_i, \mathbf{g}_i) = c_{f_A}^{(0)} \text{vec}(E_{x_i})^\dagger S_A \text{vec}(\rho) \quad (3)$$

where $\text{vec}(\cdot)$ is the vectorization operator and where S_A is now the PTM representation of the gate-sequence defined in Eq. (2). The average behavior of this sequence function correlates with how a certain outcome x_i is modified by noise, depending on how we pick A . A particularly convenient choice for the probe operator A is to set it as a subspace projector. In this way, the sequence correlation functions only retain targeted information on the gate noise. Such choice can, for example, target crosstalk noise. In general, the freedom to choose the probe operator is the key property of the gate-set-shadow method and it is this freedom that allows us to estimate different observables in the postprocessing phase (see Fig. 2 for an overview of the postprocessing stage of the protocol).

Averaging the sequence correlation functions for each sequence length m leads to the average sequence function $\hat{k}_{f_A}(m)$ that converges to an exponentially decaying function [21]. Computing the average sequence function $\hat{k}_{f_A}(m)$, for each value of m , can be achieved either by

using the empirical mean, $\hat{k}_{f_A}(m) = \frac{1}{S} \sum_{i=1}^S f_A(x_i, \mathbf{g}_i)$, or the median-of-means (MoM) given as

$$\hat{k}_{f_A}(m) = \text{median} \left[\frac{1}{N} \sum_{i=k}^{k+N-1} f_A(x_i, \mathbf{g}_i) \right], \quad (4)$$

$$k \in \{1, N+1, 2N+1, \dots, (K-1)N+1\}$$

The median-of-means is known for its greater robustness against outliers and its ability to perform better under heavy-tailed distributions [31]. This feature can be particularly useful when computing confidence intervals for finite-size datasets, as it will become clearer in the next section.

Throughout this work, we will exemplify how multiple RB protocols can be formalized under the gate-set shadow protocol, identifying instances where the protocol may ease the process of obtaining common RB metrics, and where it may actually offer additional information, otherwise impossible to extract from the original RB protocols. We refer to Appendix B for a more technical discussion on the construction of the sequence correlation function and its expectation value.

III. STANDARD AND INTERLEAVED RB

The main goal of various randomized benchmarking (RB) techniques is to characterize the quality of the experimentally implemented quantum operations. Conventionally, we quantify this quality through the average gate fidelity $\bar{F}(\mathcal{E}, U)$, where U is the desired unitary gate and $\mathcal{E}$ is its experimental implementation. In the Pauli Transfer Matrix (PTM) notation, we can define the average fidelity as [32]

$$\bar{F}(\mathcal{E}, U) = \frac{\text{Tr}(U^\dagger \mathcal{E}) + d}{d(d+1)}, \quad (5)$$

where U is the PTM representation of the ideal unitary and d is the dimension of the Hilbert space. Both standard RB [11, 12] as well as interleaved RB [14] aim to estimate this average gate fidelity. However, it is important to note that standard and interleaved RB have different goals: standard RB aims to characterize the average fidelity over a gate set, while interleaved RB aims to characterize the average fidelity of a *single gate*. Mathematically, standard RB is designed to find $\bar{F}(\Lambda, \mathcal{I})$, where $\mathcal{I}$ is the identity operator and Λ is the average noise of the full gate-set [12], while interleaved RB strives to find an estimate for $\bar{F}(\tilde{U}, U)$, where $\tilde{U}$ is the noisy implementation of U [14]. The main insight of RB is that, as long as we only aim to quantify the average noise strength, it is sufficient to estimate the trace $\text{Tr}(\dots)$ rather than reconstructing the full PTM matrix of the noise Λ . Thus, the challenge in RB is simplified to constructing efficient methods to experimentally estimate the trace.

In RB, the trace is obtained by measuring the average sequence fidelity as a function of the number of applied gates (i.e., the sequence length) [11, 12, 22]. The experimentally obtained data is then fitted to an exponential decay as a function of the sequence length m , i.e. a function of the type $b_0 + a_0 \lambda^m$. When using the Clifford gate-set, fully random sequences of gates give rise to $\lambda \propto \text{Tr}(\Lambda, \mathcal{Z})$. On the other hand, the interleaved sequences employed in interleaved RB produce a fitting model where $\lambda \propto \text{Tr}(\Lambda_U \Lambda, \mathcal{Z})$. Here, Λ_U denotes the average error modifying the experimental implementation of the gate U . While $\text{Tr}(\Lambda_U \Lambda, \mathcal{Z})$ is not directly related to the target fidelity we would like to estimate, we can obtain a bounded value for $\bar{F}(\tilde{U}, U)$ by combining standard and interleaved RB [14, 33] to obtain a single value estimate using the same strategy as in Ref. [14]. The gate-set shadow protocol can be used to emulate both standard and interleaved RB. It comes as an advantage anytime the choice of gate-set differs from the standard multi-qubit Clifford group. Character RB has a similar goal [25, 26], allowing to benchmark gate-sets beyond the standard Clifford group. The gate-set shadow protocol allows for a much more flexible and less experimentally costly way of implementing character RB. In the next section, we consider a specific example of emulating interleaved RB using a gate-set other than the Clifford group. We refer the reader to appendices B 1-B 2 for the derivation of the interleaved fitting models presented in this work.

A. Example: Interleaved RB beyond the Clifford group

While most applications of RB like protocols make use of the Clifford group as a gate-set, this might not be the most convenient choice. Particularly, if building up Clifford operations requires applying a long sequence of native gates, then the process can be demanding in terms of the required coherence time [25]. Furthermore, this issue becomes even more pressing for multiqubit systems, since the size of the Clifford group scales as $2^{\mathcal{O}(q^2)}$, with q the number of qubits [17]. Here, we consider the task of benchmarking a CNOT gate, using a smaller gate-set, G . We take G to be a particular case of the CNOT-Dihedral gate-set ($G(1)$ in [34]) which forms a group [23, 26, 34]. All the gates in this gate-set can be built from the elementary gates listed in $\mathcal{S}$ (see Appendix B 1 b for more details):

$$\mathcal{S} = \{X \otimes \mathbb{1}, \mathbb{1} \otimes X, \text{CNOT}_{0,1}, \text{CNOT}_{1,0}\}. \quad (6)$$

In total, G is formed by 24 distinct gates, i.e. the same size as the single-qubit Clifford gate-set. Assuming that all gates experience the same noise on average, averaging all implemented noisy random sequences yields an effective noise process similar to depolarizing noise. In contrast to the Clifford group, the average over random

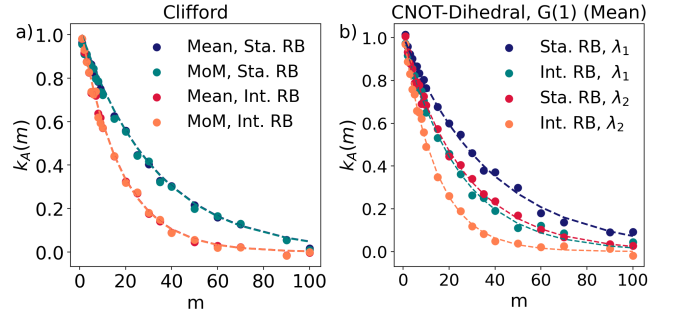


FIG. 3. Average sequence functions, $k_A(m)$, for different choices of gate-set and different choices of the random sequences. Recall that m is the length of the sequences. The function $k_A(m)$ either emulates the sequence fidelity of standard (sta.) RB (random sequences) or interleaved (int.) RB (interleaved sequences). Panel a) makes use of the Clifford gate-set, while uses the $G(1)$ CNOT-Dihedral gate-set. Panel b) shows the sequence correlation function associated with the decay rate λ_1 , and λ_2 , respectively. These are required to estimate the average fidelity (see Eq.(7)). The estimated decay rates from the fit lead to average fidelities in Tab. I. The noise model for the simulations is discussed in the main text. For panel b), only the decays obtained using the mean are showed, since in this case both estimators yield similar results.

sequences for gate-sets that do not fully scramble the full Hilbert space leads to a different type of *depolarizing-like* noise. Instead of there being only one probability event, p , of moving to a fully mixed state on the entire Hilbert space, there are now several distinct probability events, p_i , for different smaller subspaces of the full Hilbert space. Thus, the different probability events p_i lead to additional decaying factors λ_i in the RB decay model [13, 18, 22]. Averaging over random sequences produced by the concrete gate-set G in Eq.(6) gives rise to a decay profile with two decaying factors, λ_1 and λ_2 . Isolating each λ_i is possible if we *filter out* all information in the random sequences that *lives* outside the subspace where the corresponding effective mixed state $\mathbb{1}_i$ resides. This projection is accomplished by making an appropriate choice of probe operator A in postprocessing. In particular, the convenient choice of probe operator A is to set A to be a projector onto the target subspace, $A = \mathbb{P}_i$. In the PTM notation, the projectors simply correspond to block diagonal matrices, where only one of the blocks correspond to an identity matrix and all remaining diagonal blocks are zero, see Appendix B 1 b for more details.

Similarly to interleaved RB [14], an estimate for $\bar{F}(\tilde{U}, U)$ requires us to estimate both $\text{Tr}(\Lambda)$ and $\text{Tr}(U^\dagger \tilde{U} \Lambda)$, where Λ is the average gate error over the gate-set G , and $\tilde{U}$ is the noisy implementation of the CNOT gate. This requires us to estimate the two decaying factors λ_1 and λ_2 , see Fig. 3. When the sequences of gates $\mathbf{g}_i$ correspond to purely random sequences, the decay rates are related to the trace of the average error channel $\lambda_i \propto \text{Tr}(\mathbb{P}_i \Lambda)$. Similarly, the

Gate-set	Average Estimated Gate Fidelity	Ground Truth Gate Fidelity	Interleaved Fidelity Bounds
C_2	0.98 (+0.03/ - 0.03)	0.986	[0.879, 0.997]
$G(1)$	0.979 (+0.03/ - 0.04)	0.986	[0.865, 0.997]

TABLE I. Estimated average fidelity of the CNOT gate, using the estimator outlined in Ref. [14]. Estimates were produced by simulating the gate-set shadow protocol with two gate-sets: the two-qubit Clifford gate-set, C_2 , and the CNOT-Dihedral gate-set, $G = G(1)$. Results were obtained using the sample mean for both groups, since in this case no significant difference was found between the mean and the median-of-means. The error margins on the average gate fidelity estimates correspond to the uncertainty on the fit. The interleaved fidelity bounds are calculated using Eq. (B33) in Appendix. B 1 b

interleaved sequences produce decay rates where $\lambda_i \propto \text{Tr}(\mathbb{P}_i U^\dagger \mathbb{P}_i U \Lambda_U \Lambda)$. For the standard case, we set $A = \mathbb{P}_i$ to determine each decaying factor λ_i in the postprocessing step. For the interleaved sequence, we post-process using $A = \mathbb{P}_i U^\dagger \mathbb{P}_i$ with U representing the interleaved gate. Thus, we can estimate the traces as

$$\text{Tr}(\mathcal{E}) = 1 + \sum_{i=1}^2 \text{Tr}(\mathbb{P}_i) \lambda_i, \quad \lambda_i := \frac{\text{Tr}(\mathbb{P}_i A \mathbb{P}_i \mathcal{E})}{\text{Tr}(\mathbb{P}_i)}, \quad (7)$$

with either $\mathcal{E} = \Lambda$ or $\mathcal{E} = \Lambda_U \Lambda$, see also Appendix B 2 b for further details. In Fig. (3), we show the average sequence functions, $k_A(m)$, obtained using either the two-qubit Clifford group or the CNOT-Dihedral $G(1)$ group as gate-sets and with either fully random or interleaved sequences. We perform the gate-set shadow post-processing step to estimate each decaying factor separately. We note that choosing the Clifford gate-set yields essentially the same post-processing step except that we only need to estimate one decay factor. The numerical simulations use single-qubit gates that suffer from low rate incoherent noise sources (Pauli noise). Noisy CNOT gates are modeled with a coherent error of a 2-qubit zz entangling gate given by $R_{zz}(\theta)$, with θ fixed at $\theta = 0.1$ (see Appendix D 1 for further details). As discussed, the estimated decay rates λ_i are the only required information to bound the target gate fidelity and estimate the single point average gate fidelity proposed in Ref. [11]:

$$\bar{F}^{(\text{est})}(\tilde{U}, U) = 1 - \frac{(d-1)}{d} \left[1 - \frac{(d \bar{F}_{\text{sta}} - 1)}{(d \bar{F}_{\text{int}} - 1)} \right]. \quad (8)$$

Here, $\bar{F}_{\text{sta}}$ is the average gate-set fidelity estimated from averaging over fully random sequences, and $\bar{F}_{\text{int}}$ is the equivalent quantity estimated from averaging over interleaved sequences.

In Tab. I, the estimated bound and single-point estimate for the average fidelity of the CNOT gate are provided for both gate-sets. Both single-point estimates are comparable, which supports the use of a smaller gate-set.

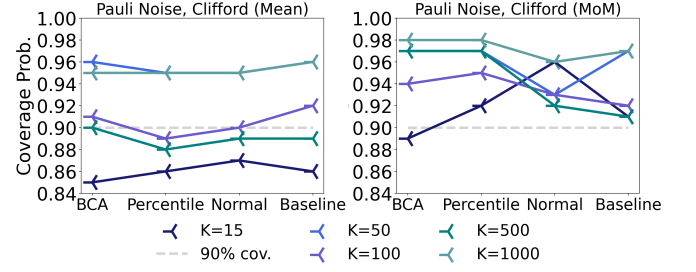


FIG. 4. Coverage probability for different confidence intervals and different sample sizes. The results are obtained by simulating the gate-set shadow protocol 100 times, and re-computing the confidence intervals, using each of the listed methods in the main text. The two plots differ by the choice of estimator: mean (left) or median-of-means (right). In this case, the protocol uses the two-qubit Clifford group as gate-set and gate errors are simulated with a Pauli noise channel.

As a final remark, we note that choosing between these two particular gate-sets also requires adapting the choice of input state and computational basis used for measurements. Indeed, the choice of the Clifford gate-set employs the z -computational basis, while the CNOT-Dihedral gate-set requires choosing the input and POVM elements as a mixed basis, where the first qubit is measured in the z -computational basis and the second qubit in the x -computational basis (see Eq.(B25) in Appendix B 1 b). While this may at first seem like a disadvantage, it is rather a consequence of the protocol's flexibility: if on the one hand it forces more conscious initial choices from the user, at the same time it grants the freedom to optimize these choices and achieve the best compromise between fidelity estimation accuracy and experimental overhead.

B. The median-of-means and robust confidence intervals

In both classical shadow tomography [20] and the gate-set shadow protocol [21], the median-of-means (MoM) plays a central role in establishing an upper bound for the number of required samples. The MoM estimator is known to be a more robust estimator compared to the empirical mean whenever outliers corrupt the sampled data or when the underlying distribution is heavy-tailed [31]. In this section, we provide further support for the MoM robustness, particularly with regards to the coverage probability of estimated confidence intervals.

Our approach is to exploit the data-driven framework of the gate-set shadow protocol to construct approximate confidence intervals using nonparametric bootstrap techniques [35, 36]. Our target parameter is the corresponding decay factor λ_i . To estimate the coverage probability, for each sample size, we simulate the gate-set shadow protocol 100 times, each time re-computing the approximate confidence intervals. The coverage probability is then defined as the fraction of times the ground truth value was

contained in the predicted interval. The target parameters λ_i are estimated using both the empirical mean and the MoM. In Fig.(4), we compare the coverage probability for the Clifford group, C_2 , for different sample sizes and different methods for estimating confidence intervals, by simulating the noisy gates using an incoherent noise source (Pauli noise) for different number of repetitions K . The gray line in Fig.(4) refers to the desired confidence level of the estimated confidence interval. For larger sample sizes, all methods provide conservative estimates for the confidence interval: all tend towards a higher coverage probability. However, decreasing the sample size leads to a coverage probability below the target confidence level for the mean estimator. Moreover, the BCa (bias-corrected and accelerated) bootstrap method tends to provide higher coverage probabilities for $K > 15$. The baseline method corresponds to bounding the estimate by a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm as often employed in practical implementations of RB protocols [37]. The MoM estimator has, generally, a better performance under smaller sample sizes, leading to a coverage probability that does not fall below the target confidence level. We refer to Appendix C for further details on the numerical estimations of the confidence intervals.

IV. SIMULTANEOUS AND CORRELATED RB

The ability to selectively target and perform operations over a sub-set of qubits in the system is crucial for any successful scalable quantum processor. Such goal requires, for instance, very precise control fields that are able to carry out target operations without unintentionally affecting the state of the idle qubits. Yet, control fields can sometimes spillover to other qubits outside of the target set and induce unintended changes. These errors are termed operation or classical crosstalk [18, 38]. Additionally, residual interactions between qubit subsystems may persist, coupling their dynamics and preventing selective operational control [38].

Within the framework of RB techniques, simultaneous RB offers a route to detect operational crosstalk [18]. The procedure works by comparing the error rates obtained from separate individual RB experiments with the error rates produced by applying the RB protocol simultaneously to all partitions of qubits [18]. In addition to operational crosstalk, it is important to know whether noise sources induce correlations between qubit subsystems since correlated crosstalk is particularly damaging for quantum error correction [3, 39, 40]. Correlated RB is a step towards distilling information on the locality and weight of crosstalk errors from simultaneous RB experiments [19]. In this section, we explore how the gate-set shadow protocol may be used to more easily extract the same information as simultaneous RB and correlated RB. Additionally, we provide examples on how the protocol

may be used to gain further insights into the presence of correlated crosstalk noise. For the sake of concreteness, we will focus on a two-qubit system.

A. Detecting operation crosstalk

To illustrate the protocol, we focus on the task of estimating the extra errors induced on the first qubit by unintentionally operating a second qubit. In other words, we use the gate-set $\mathcal{C}_1 \times \mathcal{I}$ for the case where only target operations are applied to the first qubit, and the local Clifford group $\mathcal{C}_1 \times \mathcal{C}_1$ [21] when both qubits are simultaneously targeted by single-qubit Clifford gates.

When using the gate-set $\mathcal{C}_1 \times \mathcal{I}$, we can calculate the average gate fidelity as

$$\bar{F}_1 = \frac{\text{Tr}(\mathbb{P}_1 \Lambda) + d_1}{d_1(d_1 + 1)} = \lambda_1 + \frac{1 - \lambda_1}{d_1}, \quad (9)$$

where $\lambda_1 = \frac{\text{Tr}(\mathbb{P}_1 \Lambda)}{d_1^2 - 1}$ and $d_1 = 2$. The projector $\mathbb{P}_1$ denotes the projector onto the first qubit subspace. Employing the gate-set shadow protocol to estimate Eq.(9) requires averaging over the relevant sequence correlation function $f_A(x, \mathbf{g})$. This is achieved by fixing A to be $A = \mathbb{P}_1$. Additionally, we prepare the qubit in the state $|0\rangle$ and measure in the z-computational basis, such that we have the POVMs given by

$$\begin{aligned} \text{vec}(E_0) &= \text{vec}(|00\rangle\langle 00|) + \text{vec}(|01\rangle\langle 01|), \\ \text{vec}(E_1) &= \text{vec}(|10\rangle\langle 10|) + \text{vec}(|11\rangle\langle 11|). \end{aligned} \quad (10)$$

It follows that the gate-set shadow protocol allows to retrieve a single depolarization parameter by fitting the sequence function to the model (see Appendix B3b for proof)

$$k_A(m) = c_0 \lambda_1^{m-1}. \quad (11)$$

The constant c_0 absorbs the contribution of state-preparation and measurement (SPAM) errors.

For the gate-set $\mathcal{C}_1 \times \mathcal{C}_1$, we find the effective depolarizing parameters $\lambda_{i|j}$ and λ_{ij} , where $\lambda_{i|j}$ is the effective depolarizing parameter of qubit i , while simultaneously applying random gates to subsystem j , and λ_{ij} corresponds to the effective depolarizing parameter of the joint system of i and j . The efficient extraction of these parameters is accomplished by estimating, in the postprocessing phase, the mean value of the sequence correlation function, $f_A(\mathbf{g}, A)$, over our dataset using different choices of the probe operator A summarized in Tab. II. In each of the three post-processing cases, we can fit a single exponential decay model, as in Eq. (11), to extract $\lambda_{i|j}$ and λ_{ij} .

In Fig. (5), we see that the gate-set shadow protocol, with the gate-sets $\mathcal{C}_1 \times \mathcal{I}$ and $\mathcal{C}_1 \times \mathcal{C}_1$, can be used to estimate λ_1 and $\lambda_{1|2}$. Moreover, in Fig. (5) we see that the

A	Subspace	Depolarization parameter
$\mathbb{P}_1$	Traceless subspace of qubit 1	$\lambda_{1 2} = \frac{1}{3}\text{Tr}(\mathbb{P}_1\Lambda)$
$\mathbb{P}_2$	Traceless of qubit 2	$\lambda_{2 1} = \frac{1}{3}\text{Tr}(\mathbb{P}_2\Lambda)$
$\mathbb{P}_3$	Joint traceless subspace of qubit 1 and qubit 2	$\lambda_{12} = \frac{1}{9}\text{Tr}(\mathbb{P}_3\Lambda)$

TABLE II. List of probe operators yielding different effective depolarizing parameters for the gate-set shadow protocol applied to the local Clifford group $\mathcal{C}_1 \times \mathcal{C}_1$. In each fitting model, c_i is a SPAM dependent constant, while the corresponding depolarizing parameters depend only on the average error channel Λ affecting the gates. For a more precise definition of the projectors $\mathbb{P}_i$, see Tab. IV in Appendix B3.

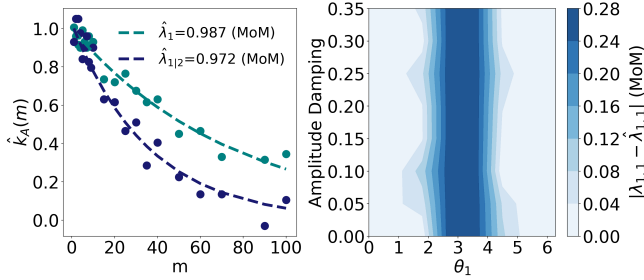


FIG. 5. Comparison of the decay curves of λ_1 and $\lambda_{1|2}$ for the gate error channel described in the main text. Each data point in the left panel results from the averaging of 1000 points, representing the different realizations of the random circuit at fixed circuit length m . The panel on the right shows the absolute difference between the estimated λ_1 and its ground truth value, based on the error model described in the main text. The estimates for the target parameters were obtained in both figures using the median-of-means (MoM).

gate set shadow protocol remains a robust estimator of λ_1 over a variety of gate error strengths. The gate errors were modeled such that gate operations on qubit-1 induce an amplitude damping channel on qubit-2 with the damping strength varied in Fig. (5). Additionally, each operation on qubit-1 was followed by a single qubit rotation $R_z(\theta_1)$. In Appendix D2, we have provided more details on the error model. The corresponding estimated values for λ_1 and $\lambda_{1|2}$, shown as an inset in the figure, are all the required information to evaluate the crosstalk metric proposed in Ref. [18]. More precisely, the error rate affecting qubit-1, when no operations are applied to the second, is simply $r_1 = 1 - \bar{F}_1$, with $\bar{F}_1$ given by Eq.(9). Additionally, the error rate affecting qubit-1 when also simultaneously applying random gates to qubit-2, $r_{1|2}$, can be calculated in the same way. Determining how the error rate of the first qubit changes due to controlling the second can be found directly in postprocessing simply by determining $\delta r_{1|2} = |r_1 - r_{1|2}| = |\lambda_1 - \lambda_{1|2}| \frac{(d_1-1)}{d_1}$.

B. Detecting correlated crosstalk

We now focus on the case where crosstalk is induced by correlated gate noise. In particular, we consider coherent crosstalk errors [19, 38]. Our goal is to identify simple ways in which the gate-set shadow protocol can be useful for extracting diagnostic metrics for correlated noise. One immediate advantage of the present protocol is in obtaining the correlated metric $\delta\lambda = \lambda_{12} - \lambda_{1|2}\lambda_{2|1}$, proposed in [18], while keeping the robustness to SPAM errors. Standard application of simultaneous RB would require taking additional measurements and combine them to effectively estimate each decaying factor individually [18]. This strategy, however, is sensitive to SPAM noise. Character RB is known to enable efficient single-decay extraction for the $\mathcal{C}_1 \times \mathcal{C}_1$ gate-set [25, 26], but can increase experimental overhead. Within the gate-set shadow protocol, the efforts to isolate a specified decaying factor are fully done in the postprocessing phase. Hence, no additional experimental overhead is incurred beyond the standard application of the protocol itself. We refer to Appendix B3 for further discussion of the fitting models corresponding to the local Clifford group as gate-set.

The central idea proposed in Ref. [21] is based on the reconstruction of the unital noise marginals of the gate noise. The unital marginals are defined as $\Lambda_i = \mathbb{P}_i\Lambda\mathbb{P}_i$ [21]. We can estimate the Λ_i from the gate-set shadow protocol with a single experimental dataset by repeating the post-processing step with the probe operator A equal to each local Clifford element C . From each of these elements we get a decay factor λ_C and we can estimate the unital marginal as

$$\Lambda_i = A_i \sum_C \lambda_C \mathbb{P}_i C^\dagger \mathbb{P}_i, \quad (12)$$

where A_i is a normalization constant, see Appendix E.

The crosstalk metric $\delta\lambda$ can be generalized to the level of the unital marginals [21]. To make this statement more concrete, we focus on a 2-qubit system and consider the following metric

$$\delta\mathcal{R}_2 = |(\Lambda_1 \otimes \Lambda_2)^{-1} \Lambda_3|. \quad (13)$$

For uncorrelated noise, $\delta\mathcal{R}_2 = \mathbb{1}$ [21]. In Fig.(6), we show $\delta\mathcal{R}_2$ for coherent gate errors modeled either by a xx , yy or zz entangling gate. The presence of correlations is visible from the higher weight matrix elements in $\delta\mathcal{R}_2$, preventing $\delta\mathcal{R}_2$ from being a perfect identity matrix. Yet, like crosstalk RB [19], we would like to characterize directly the locality and weight of crosstalk errors. To achieve this, we use the *full* unital marginal given by

$$\Lambda_F = \sum_{i=1}^n \Lambda_i, \quad (14)$$

and transform it into Pauli noise (see Eq.(E3) in Ap-

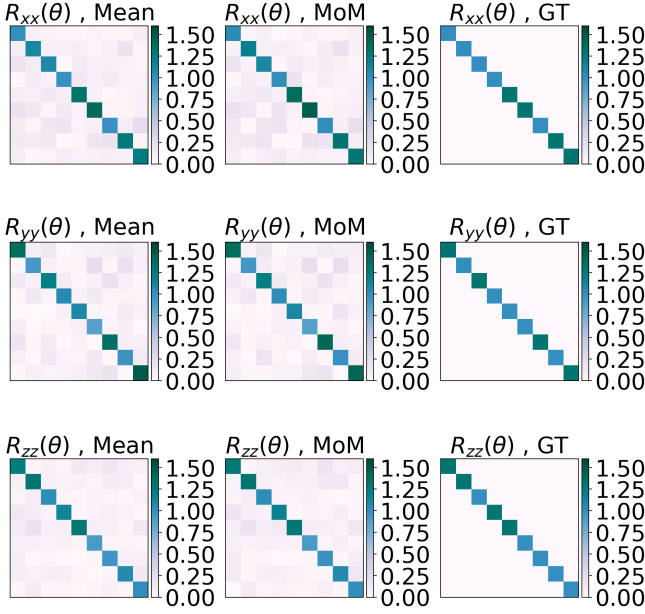


FIG. 6. $\delta\mathcal{R}_2$ for coherent gate errors, given by: $R_{xx}(\theta)$ (top panel), $R_{yy}(\theta)$ (middle panel) and $R_{zz}(\theta)$ (bottom panel). For all three cases, the angle θ was fixed at $\theta = 0.5$. In each row, from left to right: the first two figures are obtained by simulating the gate-set shadow protocol, while the third figure corresponds to the actual value of $\delta\mathcal{R}_2$, derived using full knowledge on the simulated error model (the ground truth (GT) value). The data was simulated using 15000 random sequences per sequence length.

pendix E). The goal is to find the dominant weight and locality features of the underlying gate noise from the resulting Pauli error rates.

In Fig.(7), we show the effective Pauli error rates attained in postprocessing for the correlated coherent errors (also shown in Fig. 6). In Tab. III, we show the corresponding correlated RB amplitudes, ε_i [19]. These are the coefficients defining a new parametrization of the simultaneous RB depolarizing noise channel in terms of fixed-weight depolarizing amplitudes. Note that the explicit relation between the amplitudes $\{\varepsilon\}$ and the simultaneous RB decaying factors can be found in Refs. [19, 22]. In general, these amplitudes agree with the corresponding Pauli error rates, in the sense that they are able to correctly identify the dominant weight subspace contribution. However, contrary to the noise marginals, the set of amplitudes $\{\hat{\varepsilon}\}$ cannot provide more detailed information on the errors within each subspace. A pathological example of this effect is a CNOT-like error channel. As seen in Tab. III, the error amplitude in the joint subspace ε_3 does not capture the explicitly correlated errors from the CNOT-type error channel [18]. However, in the reconstructed Pauli representation, see Fig. 8, we readily observe the CNOT-like error. We refer to Appendix E for additional details on the crosstalk metrics discussed in this section.

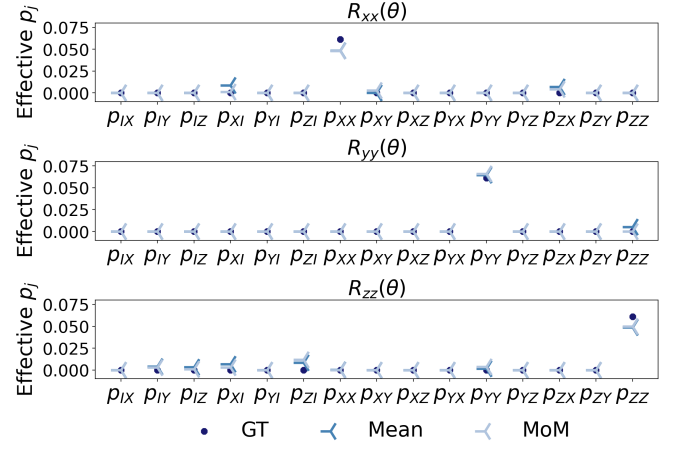


FIG. 7. Estimated Pauli error rates for the reconstructed unital marginals. The gate error model applied in the simulations corresponds to one of the following three two-qubit rotation gates: $R_{xx}(\theta)$ (top panel), $R_{yy}(\theta)$ (middle panel) and $R_{zz}(\theta)$ (bottom panel). For all three cases, the angle θ was fixed at $\theta = 0.5$. The ground-truth (GT) values are derived using full knowledge on the simulated error model.

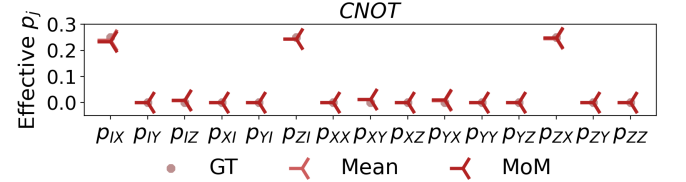


FIG. 8. Estimated Pauli error rates for the reconstructed unital marginals. The gate error channel applied in the simulations corresponds to a CNOT channel. The ground-truth (GT) values are derived using full knowledge on the simulated error model.

V. LEAKAGE RB

All the RB variants discussed so far operate under the assumption of no leakage outside the computational subspace. However, many promising quantum computing architectures, such as superconducting qubits [6] or qubits defined in quantum dots arrays [41], define their qubit states embedded in a multi-level environment [27, 40].

Leakage RB [27, 42] aims to characterize the average probability of *leaking out* of the computational subspace, referred to as leakage (L), and the average probability of the reverse process, known as seepage (S). We can define these with respect to the gate noise Λ as [27]

$$\begin{aligned} L &= \frac{1}{d_1} \text{vec}(\mathbb{1}_2)^\dagger \Lambda \text{vec}(\mathbb{1}_1), \\ S &= \frac{1}{d_2} \text{vec}(\mathbb{1}_1)^\dagger \Lambda \text{vec}(\mathbb{1}_2), \end{aligned} \quad (15)$$

where d_1 and d_2 are, respectively, the dimension of the computational and leakage subspace.

Gate Noise Model	$\hat{\varepsilon}_1$	$\hat{\varepsilon}_2$	$\hat{\varepsilon}_3$	$\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$
$R_{xx}(\theta)$	0.002 (+0.005/ - 0.002)	0 (+10 ⁻⁴)	0.068 (+0.002/ - 0.002)	{0, 0, 0.068}
$R_{yy}(\theta)$	0 (+0.002)	0 (+0.003)	0.07134 (+0.002/ - 0.003)	{0, 0, 0.068}
$R_{zz}(\theta)$	0 (+0.002)	0.002 (+0.003/ - 0.002)	0.068 (+0.003/ - 0.003)	{0, 0, 0.068}
CNOT	0.684 (+0.068/ - 0.069)	0.662 (+0.074/ - 0.075)	0 (+0.149/ - 0.281)	{2/3, 2/3, 0}

TABLE III. Estimated correlated RB amplitudes, $\{\hat{\varepsilon}\}$, for the new parametrization of the simRB depolarizing noise. The predicted amplitudes are obtained by simulating the gate-set shadow protocol with the same gate error models as in Figs.(7)-(8). The values reported were obtained using the median-of-means (MoM). The last column shows the actual values for the $\{\varepsilon\}$ coefficients for the simulated noise. The notation is as follows: ε_1 corresponds to a weight-1 coefficient, where all weight is in the qubit-1 subspace, ε_2 is a weight-1 coefficient, with all the weight in the qubit-2 subspace, and finally ε_3 is a weight-2 channel, with weight on both qubit subspaces. For more details on the evaluation of the amplitudes $\{\hat{\varepsilon}\}$, see Appendix E.

For concreteness, we consider a singlet-triplet qubit encoded in two electron spins [27]. This choice delegates the states, with total spin of $m = 1$ and $m = -1$, to the leakage subspace. We will focus on the characterization of leakage and seepage rates in the case where the target computational gates are single-qubit Pauli gates. With this aim in mind, we consider the gate-set shadow protocol implemented with the following gate-set G [27]

$$G = \langle X_C \oplus Z_L, Z_C \oplus H_L \rangle. \quad (16)$$

Here, X_C (Z_C) refers to the Pauli-X (Z) gate applied to the computational subspace, while Z_L (H_L) refers to the Pauli-Z (Hadamard) gate applied to the leakage subspace. In the definition of G , the notation $\langle \cdot \rangle$ indicates that all the elements in this gate-set can be built from the elementary gates listed inside $\langle \cdot \rangle$ (see Appendix B 4 for a description of all the elements in G). In total, G is formed by 16 different elements. Note that the gate-set G cannot induce either leakage or seepage. Thus, we are aiming to characterize leakage caused by gate noise.

For the gate-set G , there are two projectors onto the sub-spaces where the qubit states remain unchanged by all gate operations, see Appendix B 4. These projectors can be expressed as

$$\begin{aligned} \mathbb{P}_0^{(1)} &= \frac{1}{4} \text{vec}(\mathbb{1} \otimes \mathbb{1}) \text{vec}(\mathbb{1} \otimes \mathbb{1})^\dagger, \\ \mathbb{P}_0^{(2)} &= \frac{1}{4} \text{vec}(Z \otimes \mathbb{1}) \text{vec}(Z \otimes \mathbb{1})^\dagger, \end{aligned} \quad (17)$$

where the tensor product is defined in the PTM space of the triplet and singlet qubit and its leakage space. We take the input state to be initialized in the computation subspace and take the POVMs to be:

$$\text{vec}(E_x) = \{\text{vec}(\mathbb{1}_1), \text{vec}(\mathbb{1}_2)\}, \quad (18)$$

$$\text{vec}(\rho) = \text{vec}(\mathbb{1}_1). \quad (19)$$

With this choice of POVMs, and the probe operator given as

$$A = \mathbb{P}_0 := \mathbb{P}_0^{(1)} + \mathbb{P}_0^{(2)}, \quad (20)$$

we find that the sequence correlation function is

$$f_{\mathbb{P}_0}(x, \mathbf{g}) = \begin{cases} 1, & \text{if } E_x = \mathbb{1}_1 \\ 0, & \text{otherwise} \end{cases}. \quad (21)$$

In the absence of SPAM errors, this correlation function leads to the average sequence function model

$$k_{\mathbb{P}_0}(m) = \frac{S}{S+L} + \frac{L}{S+L} \lambda^m, \quad (22)$$

with $\lambda = (1 - S - L)$. This model can then be used as the fitting function, allowing the leakage and seepage rates to be estimated [27, 42], see also Appendix B 4.

The flexibility of the method allows us to pick another probe operator. Let us consider

$$A = \mathbb{P}_0^{(2)}, \quad (23)$$

which generates a sequence correlation function that alternates in sign, depending on the choice of POVM element

$$f_{\mathbb{P}_0^{(2)}}(x, \mathbf{g}) = \begin{cases} 1, & \text{if } E_x = \mathbb{1}_1 \\ -1, & \text{otherwise} \end{cases}. \quad (24)$$

When no SPAM errors are present, the resulting average sequence function reads

$$k_{\mathbb{P}_0^{(2)}}(m) = \frac{S-L}{S+L} + \frac{2L}{S+L} \lambda^m. \quad (25)$$

Both of these models may be used to extract an estimate for L and S with similar accuracy, see Fig. 9. While the leakage fitting is not SPAM insensitive, we note that if SPAM errors can be treated as a perturbation of the ideal POVM operators and input state, then the results obtained remain approximately valid.

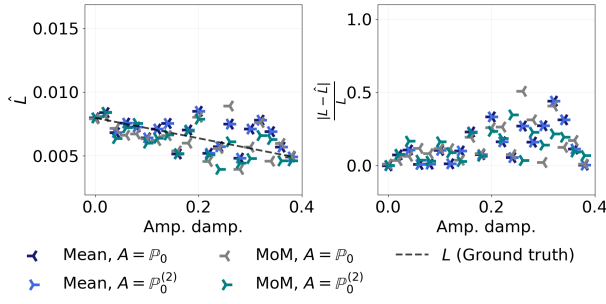


FIG. 9. Estimated leakage rate, $\hat{L}$, using as fitting model either Eq.(22) or Eq.(25). The panel on the right shows the estimated $\hat{L}$ versus the strength of the non-unital noise, here modeled by taking the tensor product of two single-qubit amplitude damping (amp. damp.) noise models, with the same amplitude damping probability. The gate noise channel employed in both figures results from the composition of Pauli noise with the resulting amplitude damping noise. The black line corresponds to the exact value of L . The left panel shows the normalized residuals between predictions and ground truth values. The data was generated by taking 1000 random sequence repetitions per sequence length.

VI. CONCLUSIONS

In this work, we have explored the gate-set shadow protocol as a unified formalism for various RB protocols and formulated concrete examples for how to use it in practical experimental circumstances. We can view each RB variant as a different realization of the gate-set shadow protocol, with different choices for: input state, POVM elements, gate-set and probability distribution defining the selection of random gates. For other choices of gate-set beyond the multi-qubit Clifford group, the resulting fitting model often harbors the contribution of many different decay factors [18, 27]. To decrease the fitting complexity of the model, character RB can be

implemented [26, 27]. Throughout the use cases explored here, we have demonstrated how the present protocol effectively allows to perform character RB at the post-processing level. This ability obliterates the need for further adaptations of the protocol at the data-collection phase. We have also identified instances where additional choices of the probe operator A give rise to additional gate noise metrics that would otherwise be impossible to obtain from the original RB protocol. The study of the confidence intervals highlights how the information extracted from the gate-set shadow protocol can be further enhanced by computational statistical methods or, more generally, by data-driven approaches. This may bring speed-up benefits that can decrease the sample overhead and lead to, for example, the reconstruction of the error marginals with less demanding sample requirements. We envision the gate-set shadow protocol as a flexible protocol that allows for benchmarking routines to be tailored and optimized for each specific experiment.

ACKNOWLEDGMENTS

This publication is part of the ‘Quantum Inspire – the Dutch Quantum Computer in the Cloud’ project (with project number [NWA.1292.19.194]) of the NWA research program ‘Research on Routes by Consortia (ORC)’, which is funded by the Netherlands Organization for Scientific Research (NWO). This work is part of the project Engineered Topological Quantum Networks (Project No.VI.Veni.212.278) of the research program NWO Talent Programme Veni Science domain 2021 which is financed by the Dutch Research Council (NWO).

DATA AND CODE AVAILABILITY

The code and data used in this work are openly available at [28].

Appendix A: Mathematical Preliminaries

1. Pauli transfer matrix formalism

Throughout this work, we explicitly made use of the Pauli transfer matrix (PTM) representation of quantum operations (also known as the Liouville representation). This representation exploits the fact that the Pauli matrices form an orthonormal basis, with respect to the Hilbert-Schmidt inner product, i.e.:

$$\frac{1}{2}\text{Tr}(P_i^\dagger P_j) = \delta_{i,j}, \quad P_i, P_j \in \{\mathbb{1}, X, Y, Z\}. \quad (\text{A1})$$

This allows us to express any linear operator, such as a single qubit density matrix ρ , as a linear combination of the Pauli matrices. More generally, the same orthonormality property also holds true for the set of normalized n -qubit Pauli matrices, defined as:

$$\frac{1}{\sqrt{2^n}} \{\mathbb{1}, X, Y, Z\}^{\otimes n} \equiv \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}^{\otimes n}. \quad (\text{A2})$$

Hence, we can use the n -qubit Pauli matrices as basis vectors, by employing the vectorization operation. This is a linear transformation that maps a matrix onto a vector. Note that since the n -qubit Pauli matrices are $d \times d$ matrices, their vectorization produces vectors with length d^2 . Thus, the vectorization process moves us from a Hilbert space of dimension d to a Hilbert space of dimension d^2 . It is common to denote the vectors in the higher-dimensional Hilbert space by employing the notation $|\cdot\rangle\rangle$ [17, 27]. Although we did not use this convention in the main text, instead explicitly highlighting the vectorization transformation $\rho \rightarrow \text{vec}(\rho)$, we adhere to it here. Using this notation, the orthogonality condition in Eq.(A1) becomes equivalent to the condition $\langle\langle\sigma_i|\sigma_j\rangle\rangle = \delta_{i,j}$, and linear operators, such as density matrices, become vectors, with the basis decomposition:

$$|\rho\rangle\rangle = \sum_{i=0}^{d^2-1} \text{Tr}(\sigma_i^\dagger \rho) |\sigma_i\rangle\rangle = \sum_{i=0}^{d^2-1} \langle\langle\sigma_i|\rho\rangle\rangle |\sigma_i\rangle\rangle, \quad (\text{A3})$$

where $d = 2^n$. Quantum operations, also known as quantum channels [43], are linear operators that map density matrices to density matrices. Any gate can serve as an example of a quantum operation (quantum channel). We will use the two terms, quantum operations and quantum channels, interchangeably. In the PTM representation, quantum channels are then linear operators that map vectors into vectors, and thus naturally allow for an explicit matrix representation. In particular, a generic quantum channel $\mathcal{E}$ can be expressed in the PTM representation as:

$$\begin{aligned} \mathcal{E}^{(\text{PTM})} &:= \mathcal{R}(\mathcal{E}) = \sum_{i,j=0}^{d^2-1} \text{Tr}(\sigma_i^\dagger \mathcal{E}(\sigma_j)) |\sigma_i\rangle\rangle \langle\langle\sigma_j| \\ &= \sum_{i,j=0}^{d^2-1} (\mathcal{R}(\mathcal{E}))_{i,j} |\sigma_i\rangle\rangle \langle\langle\sigma_j|, \end{aligned} \quad (\text{A4})$$

i.e., it is a matrix, with matrix elements given by $(\mathcal{R}(\mathcal{E}))_{i,j}$, or equivalently as $(\mathcal{R}(\mathcal{E}))_{i,j} = \text{Tr}(\sigma_i^\dagger \mathcal{E}(\sigma_j)) = \langle\langle\sigma_i|\mathcal{E}(\sigma_j)\rangle\rangle = \langle\langle\sigma_i|\mathcal{R}(\mathcal{E})|\sigma_j\rangle\rangle$. Note that in the in the main text we have assumed that all quantum channels are directly expressed in the PTM representation, that is $U \equiv U^{(\text{PTM})}$. Any sequence of transformations induced by the action of several channels on a density matrix can now be fully expressed as a sequence of matrix transformations. Thus, the composition of any two channels becomes the product of their PTM matrices, and a tensor product channel is equivalent to the tensor product of the PTM matrices of the individual channels [17, 22]:

$$|\rho'\rangle\rangle = |\mathcal{E}_b(\mathcal{E}_a(\rho))\rangle\rangle = \mathcal{R}(\mathcal{E}_b)\mathcal{R}(\mathcal{E}_a) |\rho\rangle\rangle, \quad (\text{A5})$$

$$\begin{aligned} |\rho'\rangle\rangle &= |(\mathcal{E}_b \otimes \mathcal{E}_a)(\rho_1 \otimes \rho_2)\rangle\rangle \\ &= \mathcal{R}(\mathcal{E}_b) \otimes \mathcal{R}(\mathcal{E}_a) |\rho_1 \otimes \rho_2\rangle\rangle \\ &= (\mathcal{R}(\mathcal{E}_b) |\rho_1\rangle\rangle) \otimes (\mathcal{R}(\mathcal{E}_a) |\rho_2\rangle\rangle). \end{aligned} \quad (\text{A6})$$

Quantum channels are assumed to be completely positive (CP) and trace preserving (TP) operations. These conditions guarantee that the resulting density matrix ρ' , with $\rho' = \mathcal{E}(\rho)$, is a suitable density matrix [43]: a positive operator, with trace one.

Throughout this work, we have exclusively focus on 2-qubit systems ($d = 4$). The explicit matrix representation of a generic 2-qubit quantum channel is given as:

$$\mathcal{R}(\mathcal{E}) = \begin{pmatrix} \langle\langle\tau_0| & \langle\langle\tau_{i \neq 0}| \end{pmatrix} \begin{pmatrix} 1 & \vec{0}^\top \\ \vec{w}_\mathcal{E} & R(\mathcal{E}) \end{pmatrix}, \quad (\text{A7})$$

where to make the notation more compact, we denote $|\tau_0\rangle\rangle := |\sigma_0 \otimes \sigma_0\rangle\rangle$, $\vec{0}^\top$ is the zero row vector, $R(\mathcal{E})$ is a 15×15 matrix, $\vec{w}_\mathcal{E}$ is a 15-dimensional column vector with possibly non-zero entries, and $|\tau_{i \neq 0}\rangle\rangle$ is shorthand notation for

the set of vectors $|\tau_1\rangle\rangle, |\tau_2\rangle\rangle, \dots, |\tau_{15}\rangle\rangle$, defined as:

$$\begin{aligned}
|\tau_1\rangle\rangle &:= |\sigma_0 \otimes \sigma_1\rangle\rangle, \\
|\tau_2\rangle\rangle &:= |\sigma_0 \otimes \sigma_2\rangle\rangle, \\
|\tau_3\rangle\rangle &:= |\sigma_0 \otimes \sigma_3\rangle\rangle, \\
|\tau_4\rangle\rangle &:= |\sigma_1 \otimes \sigma_0\rangle\rangle, \\
|\tau_5\rangle\rangle &:= |\sigma_2 \otimes \sigma_0\rangle\rangle, \\
|\tau_6\rangle\rangle &:= |\sigma_3 \otimes \sigma_0\rangle\rangle, \\
|\tau_7\rangle\rangle &:= |\sigma_1 \otimes \sigma_1\rangle\rangle, \\
|\tau_8\rangle\rangle &:= |\sigma_1 \otimes \sigma_2\rangle\rangle, \\
|\tau_9\rangle\rangle &:= |\sigma_1 \otimes \sigma_3\rangle\rangle, \\
|\tau_{10}\rangle\rangle &:= |\sigma_2 \otimes \sigma_1\rangle\rangle, \\
|\tau_{11}\rangle\rangle &:= |\sigma_2 \otimes \sigma_2\rangle\rangle, \\
|\tau_{12}\rangle\rangle &:= |\sigma_2 \otimes \sigma_3\rangle\rangle, \\
|\tau_{13}\rangle\rangle &:= |\sigma_3 \otimes \sigma_1\rangle\rangle, \\
|\tau_{14}\rangle\rangle &:= |\sigma_3 \otimes \sigma_2\rangle\rangle, \\
|\tau_{15}\rangle\rangle &:= |\sigma_3 \otimes \sigma_3\rangle\rangle.
\end{aligned} \tag{A8}$$

Note that we are always free to re-order the vectors $|\tau_{i \neq 0}\rangle\rangle$ and re-define their labeling, without changing the general structure of the matrix in Eq.(A7). The vector $\vec{w}_\mathcal{E}$ is formed by the matrix elements $\text{Tr}(\tau_{i \neq 0} \mathcal{E}(\tau_0)) = \frac{1}{2} \text{Tr}(\tau_{i \neq 0} \mathcal{E}(\mathbb{1}_{16 \times 16}))$. This means that only non-unital quantum channels will have $\vec{w}_\mathcal{E} \neq \vec{0}$. A unital channel is any quantum operation that leaves the identity unchanged. Any unitary gate serves as an example of a unital channel. On the other hand, a noise process like amplitude damping is not a unital operation.

2. Simple introduction to group representations in the context of gate characterization

This section provides a general introduction to the principles of finite groups and representation theory, which are used to derive analytical expressions for the fitting models in the main text. It is not intended to provide a comprehensive overview of the topic, but rather to clarify how some analytical results are obtained. Well-established results from representation theory of finite groups are introduced directly, without going over the proofs. For a more in-depth overview, see Refs. [44, 45].

The PTM formalism has allowed us to think of gates as matrix operations on a vector state. Typically gates are generated by a composition of *elementary* gate operations from a given gate-set, where the latter contains a finite number of gates. In the PTM formalism, this composition is simply a series of matrix products. We can then consider all the different matrices that can be generated by taking any combination of products of *elementary* matrices in the gate set, i.e.:

$$\text{New gate} = \prod_{g_i, g_j} \mathcal{R}(g_i) \mathcal{R}(g_j), \quad g_i, g_j \in \text{Gate-set}. \tag{A9}$$

If, through the procedure outlined in Eq.(A9), the number of distinct gates we are able to build satisfies the following requirements:

1. The number of distinct elements (matrices) that can be generated is finite. After taking a certain number of products, we merely retrieve gates that have already been generated;
2. The identity operation is contained within the set of elements we have created;
3. For each element, its inverse operation is also contained in the set of distinct elements we have produced,

then the collection of generated gates can be mathematically defined as a finite group [44], and we can identify the gate-set, G , as a matrix group. More generally, group theory allows us to study the properties of a group at an abstract level, where we may infer properties and relations between its elements without ever needing to specify concrete algebraic objects to *represent* them. In the present context, even if identifying the action of physical gates

as matrices is already an abstract construction, we are still employing a concrete basis to express these matrices, namely the PTM representation. This means we are interested in exploring the group properties of the gate-set in a concrete vector space. In very broad terms, representation theory provides the tools to study the properties of an abstract group as a matrix group of squared matrices, inheriting the same structure as the original abstract group [44]. Concretely, let g_a, g_b and g_c be elements of a group, related as $g_c = g_a \times g_b$, and $M(g_a), M(g_b)$ and $M(g_c)$ are square matrices. These matrices form an appropriate representation of the group if:

$$M(g_a)M(g_b) = M(g_c) , \text{ for all } g_a, g_b, g_c \in G . \quad (\text{A10})$$

Thinking in terms of quantum gates, this means that, whenever the gate-set G forms a group, the PTMs of its gate operations form an actual matrix representation of the corresponding group. The PTM formalism implies a particular basis choice (the Pauli basis). However, Eq.(A10) does not imply there is a unique set of matrices forming an appropriate representation of a group. Indeed, any matrix similarity transformation acting over all matrices M (such as a change of basis), $\tilde{M}(g) = SM(g)S^{-1}$, is an equally suitable representation, i.e. it likewise satisfies Eq.(A10). Since there are many possible matrix representations available, the study of how the group matrices relate to each other and how they transform vectors in their corresponding vector space cannot be dependent on our particular basis choice, but rather needs to be rooted in properties that remain invariant to it. Given the freedom in the choice of basis, we are always at liberty to find S such that $SM(g)S^{-1}$, for all $M(g)$, renders the matrices to have a block diagonal form:

$$M(g) = \begin{pmatrix} M_1(g) & \vec{0} & \dots & \vec{0} \\ \vec{0} & M_2(g) & \dots & \vec{0} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{0} & \vec{0} & \dots & M_k(g) \end{pmatrix} , \text{ for all } g \in G . \quad (\text{A11})$$

Each matrix $M_i(g)$ in Eq.(A11) is a lower-dimensional matrix. Its dimensionality can even be as low as a 1×1 matrix, giving rise to a scalar. In representation theory, we search for an S that can simultaneously transform all matrices $M(g)$ into block diagonal form, while at the same time guaranteeing that this is the similarity transformation that brings them as close as possible to full diagonal form. Note that for most purposes we never need to explicitly find the transformation S . Its existence is sufficient for us to directly think of the matrix representations of the group as matrices with the structure in Eq.(A11). Recalling that $M(g)$ is a matrix defined in some vector space V , what Eq.(A11) clarifies is that a group will naturally give rise to a decomposition of the underlying space, V , in terms of smaller subspaces V_i . For each vector contained in a subspace, $\vec{v}_i \in V_i$, the action of any group element over $\vec{v}_i$ is fully dictated by the action of the lower-dimensional matrix $M_i(g)$. Since there are no non-zero matrix elements outside the diagonals, there is no way for the group operations to transform $\vec{v}_i$ into a vector belonging to a different subspace. Thus, $M(g) \vec{v} = \sum_i M_i(g) \vec{v}_i$. Consequently, to understand how the group operations induce transformations on the basis vectors, we only need to study how each lower-dimensional matrix transforms the vectors in the corresponding subspaces. In the present context, this is the precursor for the emergence of *smaller* subspaces that will render the description of the average gate noise as a type of depolarizing noise, with possibly distinct depolarizing parameters. The matrices $M(g)$ are called reducible representations, and the lower dimensional matrices, $M_i(g)$, are called irreducible representations (irreps). The irreps can be equivalent or inequivalent. Two irreducible representations are said to be equivalent if there is a similarity transformation that can relate the two, $M_j = \tilde{S}M_i\tilde{S}^{-1}$ (with $i \neq j$). In the context of finite groups, we distinguish between different irreps by their trace: inequivalent irreps have different traces, equivalent irreps share the same trace. The trace of a (matrix) representation is often referred to as the character of the representation:

$$\begin{aligned} \text{Character of the } i\text{th irrep} &:= \chi_i(g) \\ &= \text{Tr}(M_i(g)) , \end{aligned} \quad (\text{A12})$$

$$\begin{aligned} \text{Character of a reducible rep.} &:= \chi(g) \\ &= \text{Tr}(M(g)) . \end{aligned} \quad (\text{A13})$$

There are groups that, upon arriving to Eq.(A11), result in a description in terms of irreducible representations that are all different from one another (all irreps are inequivalent). These groups are called multiplicity-free groups. The n -qubit Clifford group (C_n), the local Clifford group ($C_1^{\times n}$), and the CNOT-dihedral group presented in the main text are all examples of multiplicity-free groups. On the other hand, some groups can lead to a decomposition that contains

equivalent irreps, and are called non-multiplicity-free groups. The group $\mathbb{C}_1 \times \mathcal{I}$, used in the study of operational crosstalk in Section IV A, is an example of a non-multiplicity-free group. An important result in representation theory of finite groups is that the characters of irreducible representations satisfy the orthonormal condition [44, 45]:

$$\frac{1}{|G|} \sum_{g \in G} \chi_i^*(g) \chi_j(g) = \delta_{i,j} , \quad (\text{A14})$$

where $|G|$ stands for the order of the group (total number of distinct elements in G). The number of times the same irrep appears in Eq.(A11) is referred to as the multiplicity of the irrep. The fact that inequivalent irreps lead to different characters satisfying the orthogonality condition in Eq.(A14) enables us to determine the multiplicity of an irrep with only knowledge on the character of the representations [45]:

$$\text{Multiplicity of the } i\text{th irrep} := a_i = \frac{1}{|G|} \sum_{g \in G} \chi^*(g) \chi_i(g) . \quad (\text{A15})$$

Whether or not a gate set is equivalent to a multiplicity-free group is a relevant piece of information for any quantity resulting from averaging over group operations. This will soon become clearer when considering the average gate noise.

For any finite group, there is always one irreducible representation that assumes the form of an identity operation for all the elements in the group. In Eq.(A11), this irrep will appear as a scalar in the diagonal equal to one, i.e. $M_i(g) = 1$, for all $g \in G$. Since this irrep is the same for all elements in the group and will induce no transformations over the basis vectors in V , it is often called the *trivial* irrep. In contrast, all other irreps can induce different transformations on the basis of states, depending on which element g we are considering. For this reason, they will sometimes be refer to as non-trivial irreps. The trivial irrep coincides with its character. This is true for any scalar irreducible representation. Note that since the irreps of the group give rise to a decomposition of the vector space into subspaces, the trivial irrep is associated to an one-dimensional subspace. To convey the relation between subspaces and irreps, we refer to the subspaces as simply irreducible subspaces, with the same labeling as the one ascribed to the irreps.

Weighted averages over group operations can be used to create projectors onto irreducible subspaces [26, 45]. In particular, averaging over group representations yields a projector onto the trivial subspace [45]:

$$\mathbb{P}_{\text{triv}} = \frac{1}{|G|} \sum_{g \in G} M(g) . \quad (\text{A16})$$

It is important to note that if the trivial irrep is a not multiplicity free irrep, then the projector $\mathbb{P}_{\text{triv}}$ becomes higher-dimensional, instead of a simple one-dimensional projector. Projectors to other irreducible subspaces can also be obtain by a similar averaging procedure. The main difference compared to Eq.(A16) is that the representations are weighted by the irrep character of the target subspace, i.e. $\chi_i(g)$. We will not make use of this result explicitly here, but note that it is the basis of why character RB is capable of extracting each decaying rate separately, whenever the corresponding group is multiplicity-free [26].

Taking the Kronecker product between representations provides yet another possible way of representing the group elements. If $M_i(g)$ and $M_j(g)$ are two real and irreducible representations, then the result of averaging $M_i(g) \otimes M_j(g)$ over all group elements is zero, unless $M_i(g)$ and $M_j(g)$ are equivalent irreps. This result will be extensively used over the next sections. We can also consider what is the result of averaging an arbitrary constant matrix over all group operations. In that case, instead of a single projector, we get a weighted sum of all irreducible subspace projectors [18]:

$$\begin{aligned} \text{Multiplicity free group: } \frac{1}{|G|} \sum_{g \in G} M(g) B M^{-1}(g) &= \\ &= \sum_{i=1}^n \frac{\text{Tr}(B \mathbb{P}_i)}{\text{Tr}(\mathbb{P}_i)} \mathbb{P}_i \\ &= \sum_{i=1}^n \lambda_i(B) \mathbb{P}_i , \end{aligned} \quad (\text{A17})$$

$$\begin{aligned}
\text{Non-multiplicity free group: } & \frac{1}{|G|} \sum_{g \in G} M(g) B M^{-1}(g) = \\
& = \sum_{i=1}^n \sum_{k,k'=1}^{a_i} \frac{\text{Tr}(B (\mathbb{P}_i^{(k,k')})^\top)}{\text{Tr}((\mathbb{P}_i^{(k,k')})^\top \mathbb{P}_i^{(k,k')})} \mathbb{P}_i^{(k,k')} \\
& = \sum_{i=1}^n \sum_{k,k'=1}^{a_i} \lambda_i^{(k,k')}(B) \mathbb{P}_i^{(k,k')} .
\end{aligned} \tag{A18}$$

In Eqs.(A17) and (A18), n denotes the number of distinct irreps, a_i is the multiplicity of the i th-irrep, $\mathbb{P}_i$ is the projector onto the i th irreducible subspace, $\mathbb{P}_i^{(k,k')}$ is a projector connecting two sub-spaces: it projects from the subspace associated to copy k' of i th-irrep to the subspace corresponding to copy k of the i th-irrep. The symbol $\top$ denotes matrix transposition. In the context of the present work, we note that the RB procedure of generating random sequence of gates and then averaging over them is directly related with the group averages in Eq.(A17) (if the gate-set is multiplicity free) or Eq.(A18) (if the gate-set is non-multiplicity free). This relationship exists because RB assumes that all gates experience the same noise, or at least that deviations from this limit are weak. Indeed, identifying the constant matrix B as the gate noise $\Lambda^{(PTM)}$ makes it clear that the coefficients λ_i (or $\lambda_i^{(k,k')}$) are precisely the RB decay rates. This also clarifies why certain gate-sets can give rise to fitting models with different complexity. The number of decay rates contributing to the fitting model depends on how the group decomposes into irreps, which in turn constrains the dynamic of states to smaller subspaces of the full vector space.

Appendix B: The mathematical construction of the fitting models

In this section, we provide the derivation of the fitting models presented in the main text. We note that the general procedure for constructing the gate-set shadow models is provided in Ref. [21]. To provide more clarity in the derivation of the fitting models presented here, we revisit in this section the essential steps in the general formalism derived in Ref. [21].

We begin by rewriting the correlation function, $f_A(x, \mathbf{g})$, using the PTM notation introduced in Section A 1 and the machinery of representation theory from Section A 2:

$$f_A(x, \mathbf{g}) = c_{f_A}^{(0)} \langle \langle E_x | \mathcal{R}_i(g_m) \prod_{j=1}^{m-1} A \mathcal{R}_i(g_j) | \rho \rangle \rangle . \tag{B1}$$

In Eq.(B1), g_j is a gate from a gate-set G . Its corresponding PTM matrix is given by $\mathcal{R}(g_j)$. Since we take the gate-set to form a group, each matrix $\mathcal{R}(g_j)$ can be identified as a block-diagonal matrix, with each block corresponding to an irrep matrix. Then $\mathcal{R}_i(g_j)$ is the i th-irrep block. Note that we did not specify in the main text that the sequence correlation function is derived from selecting an irrep rather than taking the full reducible PTM matrix. This omission was made in order to maintain focus on the main steps of the protocol. We start by assuming fully random sequences of gates, and discuss interleaved sequences separately. The remaining terms in Eq.(B1) are A , the user-defined probe-operator (already in the PTM formalism), and $c_{f_A}^{(0)}$ a the normalization constant. The sequence function, $k_A(m)$, specifies the fitting model, and is defined as the expectation value of $f_A(x, \mathbf{g})$. Viewing $f_A(x, \mathbf{g})$ as a random variable, evaluating its expectation value requires assigning a probability distribution function, $p(x, \mathbf{g})$, to it:

$$p(x, \mathbf{g}) = \langle \langle \tilde{E}_x | \prod_{j=1}^m \mathcal{R}(\Lambda) \mathcal{R}(g_j) | \tilde{\rho} \rangle \rangle . \tag{B2}$$

The form of $p(x, \mathbf{g})$ merits a few observations. Firstly, we note that $p(x, \mathbf{g})$ is defined as the probability that, in a noisy experimental setup and given a sequence of gates $\mathbf{g}$, the measured outcome is x . As both gate noise and SPAM noise can affect the distribution of measurement outcomes, $p(x, \mathbf{g})$ incorporates the effects of both types of noise. This is accounted for by letting $\tilde{E}_x$ and $\tilde{\rho}$ be the noise implementation of the POVM element and input state, respectively. Secondly, $p(x, \mathbf{g})$ is define through the more general reducible (PTM) representation of the gates. Lastly, we adopt the usual RB convention for modeling the noisy gates, in which the noisy implementation of an ideal gate $\mathcal{R}(g)$ is given by $\mathcal{R}(\Lambda) \mathcal{R}(g)$. This is less general than taking, as in Ref. [21], the noisy gates to be $\mathcal{R}(\Lambda_L) \mathcal{R}(g) \mathcal{R}(\Lambda_R)$. It, nevertheless, provides a more direct mapping to the RB fitting models. The sequence function is then given by the

expectation value [21]:

$$\begin{aligned}
k_A(m) &= \mathbb{E}_{\mathbf{g} \in G^{\times m}} \sum_{x \in \{0,1\}^n} f_A(x, \mathbf{g}) p(x, \mathbf{g}) \\
&= c_{f_A}^{(0)} \mathbb{E}_{\mathbf{g} \in G^{\times m}} \sum_{x \in \{0,1\}^n} \langle \langle E_x | \mathcal{R}_i(g_m) \prod_{j=1}^{m-1} A \mathcal{R}_i(g_j) | \rho \rangle \rangle \langle \langle \tilde{E}_x | \mathcal{R}(\Lambda) \mathcal{R}(g_m) \prod_{j=1}^{m-1} \mathcal{R}(\Lambda) \mathcal{R}(g_j) | \tilde{\rho} \rangle \rangle \\
&= c_{f_A}^{(0)} \mathbb{E}_{\mathbf{g} \in G^{\times m}} \sum_{x \in \{0,1\}^n} \text{Tr} \left[|\rho\rangle \langle \langle E_x | \left(\mathcal{R}_i(g_m) \prod_{j=1}^{m-1} A \mathcal{R}_i(g_j) \right) \right] \text{Tr} \left[|\tilde{\rho}\rangle \langle \langle \Lambda^*(\tilde{E}_x) | \left(\mathcal{R}(g_m) \prod_{j=1}^{m-1} \mathcal{R}(\Lambda) \mathcal{R}(g_j) \right) \right] \quad (\text{B3}) \\
&= c_{f_A}^{(0)} \mathbb{E}_{\mathbf{g} \in G^{\times m}} \sum_{x \in \{0,1\}^n} \text{Tr} \left[|\rho \otimes \tilde{\rho}\rangle \langle \langle E_x \otimes \Lambda^*(\tilde{E}_x) | \left(\mathcal{R}_i(g_m) \otimes \mathcal{R}(g_m) \right) \prod_{j=1}^{m-1} (A \otimes \mathcal{R}(\Lambda)) \left(\mathcal{R}_i(g_j) \otimes \mathcal{R}(g_j) \right) \right] \\
&= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^n} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \mathbb{E}_{g_m \in G} \left(\mathcal{R}_i(g_m) \otimes \mathcal{R}(g_m) \right) \prod_{j=1}^{m-1} (A \otimes \mathcal{R}(\Lambda)) \mathbb{E}_{g_j \in G} \left(\mathcal{R}_i(g_j) \otimes \mathcal{R}(g_j) \right) \right],
\end{aligned}$$

where we have defined $\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) := |\rho \otimes \tilde{\rho}\rangle \langle \langle E_x \otimes \Lambda^*(\tilde{E}_x) |$. As discussed in Section A 2, taking averages over representations gives rise to a projector. In particular, note that for representations [45]:

$$\mathcal{R}_i(g) \otimes \mathcal{R}(g) = (\mathcal{R}_i(g) \otimes \mathcal{R}_i^{\oplus a_i}(g)) \oplus_{j \neq i} (\mathcal{R}_i(g) \otimes \mathcal{R}_j^{\oplus a_j}(g)). \quad (\text{B4})$$

Since $\mathcal{R}_i(g)$ and $\mathcal{R}_j(g)$ are distinct irreps, only the term $\mathcal{R}_i(g) \otimes \mathcal{R}_i^{\oplus a_i}(g)$ provides a non-zero contribution when averaging over all group elements [21]. As discussed in the previous section, uniformly averaging over a representation gives rise to a projector onto the trivial subspace. This allows us write:

$$\mathbb{E}_{g \in G} (\mathcal{R}_i(g) \otimes \mathcal{R}(g)) = \mathbb{P}_{\text{triv}}. \quad (\text{B5})$$

Providing a basis decomposition for the trivial projector requires specifying the group and the choice of irrep in the sequence correlation function. Keeping the discussion general for now, Eq.(B3) can be further simplify to:

$$k_A(m) = c_{f_A}^{(0)} \sum_{x \in \{0,1\}^n} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\mathbb{P}_{\text{triv}}(A \otimes \mathcal{R}(\Lambda)) \mathbb{P}_{\text{triv}} \right)^{m-1} \right]. \quad (\text{B6})$$

1. Standard RB

a. Multi-qubit Clifford group

The multi-qubit Clifford group is a multiplicity-free group, with only two irreducible representations [21, 46]. This statement applies for any n -qubit system, but in order to simplify the discussion, we focus on a 2-qubit system. The emergence of the two distinct irreps can be made more evident by recalling the following statements about the PTM formalism. The first is the observation that we now work in a vector space spanned by the vectorized set of all 2-qubit Pauli operators. The second is that any CPTP operation in this space has the structure of the matrix in Eq.(A7). Furthermore, with the exception of the addition of a phase factor, the Clifford group acts as a permutation of the normalized Pauli vectors (see appendices of Refs. [17, 18]), i.e.: $\mathcal{R}(g)|\tau_i\rangle\rangle = \pm|\tau_l\rangle\rangle$, $g \in \mathbb{C}_n$. The exception to this permutation of Pauli vectors is the vectorized identity, that always gets mapped to itself by the action of the group. It then follows that $\mathcal{R}_{i,0}(g) = 0$, which simplifies Eq.(A7) to:

$$\mathcal{R}(g) = \begin{matrix} \langle \langle \tau_0 | \\ \langle \langle \tau_{i \neq 0} | \end{matrix} \begin{pmatrix} |\tau_0\rangle\rangle & |\tau_{i \neq 0}\rangle\rangle \\ 1 & \vec{0}^T \\ \vec{0} & R(g) \end{pmatrix}. \quad (\text{B7})$$

From the discussion in Section A 2, we recognize $\mathcal{R}(g)$ in Eq.(B7) as a reducible representation of the group. There are then two irreps: a scalar irrep (the trivial irrep), associated to the one-dimensional vector space spanned by the single vector $|\tau_0\rangle$, and a non-trivial irrep, operating on the subspace spanned by the set of vectors $\{|\tau_i\rangle\}_{i \neq 0}$. Recall that the sequence correlation function in Eq.(B1) relies in specifying a particular irrep. To then recover the standard RB fidelity from the gate-set shadow protocol, the following settings should be applied: the gate-set is fixed to be the multi-qubit Clifford group; the random sequences are constructed by uniformly sampling the group; $f_A(x, \mathbf{g})$ is evaluated using only the non-trivial irreducible representation; and A is fixed as the projector onto the corresponding non-trivial subspace. Using the basis vectors specified in Eq.(A8), the projector onto the non-trivial subspace is given as:

$$\mathbb{P}_1 = \sum_{i=1}^{15} |\tau_i\rangle\langle\tau_i|. \quad (\text{B8})$$

The trivial projector in Eq.(B5) is now evaluated with $\mathcal{R}_i(g)$ fixed as the non-trivial irrep, $\mathcal{R}_i(g) = R(g)$. Since $R(g)$ operates in the subspace spanned by the vectors $\{|\tau_i\rangle\}_{i \neq 0}$, the matrix representation $R^{\otimes 2}(g)$ has a natural decomposition in terms of the basis vectors $\{|\tau_i\rangle \otimes |\tau_j\rangle\}_{i,j=1}^{15}$. Considering that $R(g)$ permutes, with phases $\pm$, the basis vectors $\{|\tau_i\rangle\}_{i \neq 0}$, an invariant vector under the action of $R^{\otimes 2}(g)$, for any Clifford element g , is given by a uniform superposition of the vectors $\{|\tau_i\rangle \otimes |\tau_i\rangle\}_{i \neq 0}^{15}$, i.e.:

$$\vec{v}_{\text{triv}} = \frac{1}{|\mathbb{P}_1|^{1/2}} \sum_i |\tau_i\rangle \otimes |\tau_i\rangle = \frac{1}{|\mathbb{P}_1|^{1/2}} \sum_i \text{vec}(|\tau_i\rangle\langle\tau_i|) = \frac{1}{|\mathbb{P}_1|^{1/2}} |\mathbb{P}_1\rangle. \quad (\text{B9})$$

Note that we have used the relation between the Kronecker and outer product: $|i\rangle \otimes |j\rangle = \text{vec}(|i\rangle\langle j|)$ [47]. In the equation above, the constant $|\mathbb{P}_1| = \text{Tr}(\mathbb{P}_1)$ is introduced as a normalization factor. This allows us to express the trivial projector in Eq.(B5) as:

$$\mathbb{P}_{\text{triv}} = \vec{v}_{\text{triv}} \vec{v}_{\text{triv}}^\dagger = \frac{1}{|\mathbb{P}_1|} |\mathbb{P}_1\rangle\langle\mathbb{P}_1|. \quad (\text{B10})$$

The fitting model in Eq.(B6) then reduces to:

$$\begin{aligned} k_A(m) &= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\frac{1}{|\mathbb{P}_1|^2} \langle\langle \mathbb{P}_1 | (A \otimes \mathcal{R}(\Lambda)) | \mathbb{P}_1 \rangle\rangle |\mathbb{P}_1\rangle\langle\mathbb{P}_1| \right)^{m-1} \right] \\ &= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\frac{1}{|\mathbb{P}_1|} \langle\langle \mathbb{P}_1 | \mathcal{R}(\Lambda) \mathbb{P}_1 A^\top \rangle\rangle \mathbb{P}_{\text{triv}} \right)^{m-1} \right] \\ &= c_{f_A}^{(0)} \lambda^{m-1}(A^\top, \Lambda) \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_1|} \langle\langle \mathbb{P}_1 | \Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) | \mathbb{P}_1 \rangle\rangle, \end{aligned} \quad (\text{B11})$$

with $\lambda(A^\top, \Lambda) = \frac{1}{|\mathbb{P}_1|} \text{Tr}(\mathbb{P}_1 A^\top \mathbb{P}_1 \mathcal{R}(\Lambda))$. Note that in Eq.(B11), we used the matrix vectorization result for the triple matrix product ABC [48]:

$$|ABC\rangle = (C^\top \otimes A)|B\rangle. \quad (\text{B12})$$

Letting $A = \mathbb{P}_1$, the resulting decay rate can be directly related to the average gate-set fidelity (see Eq.(5)):

$$\lambda(\mathbb{P}_1, \Lambda) = \frac{1}{|\mathbb{P}_1|} \text{Tr}(\mathbb{P}_1 \mathcal{R}(\Lambda)) = \frac{d\bar{F}(\Lambda, \mathcal{I}) - 1}{d - 1}, \quad (\text{B13})$$

with d the dimension of the Hilbert space. The decay rate $\lambda(\mathbb{P}_1, \Lambda)$ is the typical figure of merit measured in standard RB. We note that, in this case, a convenient choice of POVM operators and input state is to set them as operators

defined in the z -computational basis, i.e.:

$$\begin{aligned}
|\rho\rangle &= |00\rangle = \frac{1}{2} \left(|\tau_0\rangle + |\tau_3\rangle + |\tau_6\rangle + |\tau_{15}\rangle \right) = |E_{00}\rangle, \\
|E_{01}\rangle &= \frac{1}{2} \left(|\tau_0\rangle - |\tau_3\rangle + |\tau_6\rangle - |\tau_{15}\rangle \right) \\
|E_{10}\rangle &= \frac{1}{2} \left(|\tau_0\rangle + |\tau_3\rangle - |\tau_6\rangle - |\tau_{15}\rangle \right) \\
|E_{11}\rangle &= \frac{1}{2} \left(|\tau_0\rangle - |\tau_3\rangle - |\tau_6\rangle + |\tau_{15}\rangle \right).
\end{aligned} \tag{B14}$$

It only remains to determine the normalization constant $c_{f_A}^{(0)}$. This constant is defined such that, in the absence of SPAM and gate noise, $k_A(1) = 1$. In this regime, Eq.(B11) reads:

$$\begin{aligned}
k_A(1) &= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_1|} \langle \langle \mathbb{P}_1 | \Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) | \mathbb{P}_1 \rangle \rangle \\
&= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_1|} \sum_{i \neq 0} \langle \langle \tau_i | \rho \rangle \rangle^2 \sum_{j \neq 0} \langle \langle E_x | \tau_j \rangle \rangle^2 \\
&= \frac{3}{4} c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_1|} \sum_{j \neq 0} \langle \langle E_x | \tau_j \rangle \rangle^2 = \frac{3^2}{4} c_{f_A}^{(0)} \frac{1}{|\mathbb{P}_1|} \implies c_{f_A}^{(0)} = \frac{20}{3}.
\end{aligned} \tag{B15}$$

The value of the normalization constant $c_{f_A}^{(0)}$ is in agreement with numerical simulations of the protocol, guaranteeing $k_A(1) = 1$ in the noiseless regime.

b. CNOT-Dihedral group $G(1)$

The CNOT-Dihedral group $G(1)$ is a multiplicity-free group. It is generated by the elements in the set $\mathcal{S}$, up to a global phase factor w (here $\mathbb{C}$ denotes the space of complex numbers). This means that all elements in the group can be obtained as a finite combination of elements in $\mathcal{S}$. Additionally, two elements are considered identical if they only differ by a scalar multiplicative phase factor.

$$G(1) = \langle \mathcal{S} \rangle / \langle w \mathbb{1} : w \in \mathbb{C} \rangle, \quad \mathcal{S} = \{X \otimes \mathbb{1}, \mathbb{1} \otimes X, \text{CNOT}_{0,1}, \text{CNOT}_{1,0}\}. \tag{B16}$$

This group has 24 distinct elements, and is a subgroup of the Clifford group. To construct all its elements, we make use of the algorithm proposed in Ref. [34]. This allows us to describe all elements in the group as the union of the four sets:

$$\mathcal{S}_1 = \{\mathbb{1} \otimes \mathbb{1}, \mathbb{1} \otimes X, X \otimes \mathbb{1}, X \otimes X\} \equiv \{\mathbb{1}_{d \times d}, X_2, X_1, X \otimes X\}, \tag{B17}$$

$$\begin{aligned}
\mathcal{S}_2 = \{ &\text{CNOT}_{0,1}, \text{CNOT}_{1,0}, X_1 \text{CNOT}_{0,1}, X_1 \text{CNOT}_{1,0}, \\
&X_2 \text{CNOT}_{0,1}, X_2 \text{CNOT}_{1,0}, (X \otimes X) \text{CNOT}_{0,1}, (X \otimes X) \text{CNOT}_{1,0} \},
\end{aligned} \tag{B18}$$

$$\mathcal{S}_3 = \{\text{CNOT}_{0,1} \text{CNOT}_{1,0}, \text{CNOT}_{1,0} \text{CNOT}_{0,1}, X_1 \text{CNOT}_{0,1} \text{CNOT}_{1,0}, \tag{B19}$$

$$\begin{aligned}
&X_1 \text{CNOT}_{1,0} \text{CNOT}_{0,1}, X_2 \text{CNOT}_{0,1} \text{CNOT}_{1,0}, X_2 \text{CNOT}_{1,0} \text{CNOT}_{0,1} \\
&(X \otimes X) \text{CNOT}_{0,1} \text{CNOT}_{1,0}, (X \otimes X) \text{CNOT}_{1,0} \text{CNOT}_{0,1} \},
\end{aligned} \tag{B20}$$

$$\begin{aligned}
\mathcal{S}_4 = \{ &\text{CNOT}_{0,1} \text{CNOT}_{1,0} \text{CNOT}_{0,1}, X_1 \text{CNOT}_{0,1} \text{CNOT}_{1,0} \text{CNOT}_{0,1}, \\
&X_2 \text{CNOT}_{0,1} \text{CNOT}_{1,0} \text{CNOT}_{0,1}, (X \otimes X) \text{CNOT}_{0,1} \text{CNOT}_{1,0} \text{CNOT}_{0,1} \}.
\end{aligned} \tag{B21}$$

Note that in Eq.(B16) we are representing the group elements as operations acting on a 2-qubit Hilbert space, and move to the PTM description later on. It is clear that, up to a phase, the elements in $\mathcal{S}_1$ always transform any

2-qubit Pauli operator back to itself. The action of the CNOT gates over the Pauli set $\{Z \otimes \mathbb{1}, \mathbb{1} \otimes Z, Z \otimes Z\}$ simply interchanges the Pauli operators within the set:

$$\begin{aligned} \text{CNOT}_{0,1} \{Z \otimes \mathbb{1}, \mathbb{1} \otimes Z, Z \otimes Z\} \text{CNOT}_{0,1} &= \{Z \otimes Z, \mathbb{1} \otimes Z, \mathbb{1} \otimes Z\}, \\ \text{CNOT}_{1,0} \{Z \otimes \mathbb{1}, \mathbb{1} \otimes Z, Z \otimes Z\} \text{CNOT}_{1,0} &= \{Z \otimes Z, Z \otimes \mathbb{1}, \mathbb{1} \otimes Z\}. \end{aligned} \quad (\text{B22})$$

Since the elements within the set of gates from $\mathcal{S}_2$ to $\mathcal{S}_4$ are a composition of the gates in $\mathcal{S}_1$ and the CNOT gates, the action of all group elements over $\{Z \otimes \mathbb{1}, \mathbb{1} \otimes Z, Z \otimes Z\}$ simply produces the same set, with re-ordering of the three Pauli operators. Thus, once in the PTM formalism, the vectors $\{|\tau_3\rangle\rangle, |\tau_6\rangle\rangle, |\tau_{15}\rangle\rangle\}$ form an invariant subspace. All the group elements will map the Pauli identity to itself. This enables the trivial irrep to be recovered again in relation to the one-dimensional vector space spanned by $\{|\tau_0\rangle\rangle\}$. We can proceed similarly to Eq.(B22) to obtain the action of the group elements over the remaining 2-qubit Pauli operators. This allows us to conclude that the subspace spanned by $\{|\tau_i\rangle\rangle\}_{i=\{1,2,\dots,14\}\setminus\{3,6\}}$ is also an invariant subspace. The three subspaces are disjoint and together span the full PTM vector space. Hence, the CNOT-Dihedral group $G(1)$ has three distinct irreducible representations, with dimensions 1 (trivial), 3 and 12. Identifying the irreps is important for benchmarking the fidelity of the gate-set. The standard RB figure of merit, $\bar{F}(\Lambda, \mathcal{I})$, can be decomposed in terms of the projectors onto the two non-trivial irreducible subspaces of $G(1)$:

$$\begin{aligned} \bar{F}(\Lambda, \mathcal{I}) &= \frac{1}{d} + \frac{1}{d(d+1)} \left(\text{Tr}(\mathbb{P}_1 \mathcal{R}(\Lambda)) + \text{Tr}(\mathbb{P}_2 \mathcal{R}(\Lambda)) \right) \\ &= \frac{1}{d} + \frac{1}{d(d+1)} \left(|\mathbb{P}_1| \lambda(\mathbb{P}_1, \Lambda) + |\mathbb{P}_2| \lambda(\mathbb{P}_2, \Lambda) \right), \end{aligned} \quad (\text{B23})$$

with

$$\mathbb{P}_1 = \sum_{i=\{3,6,15\}} |\tau_i\rangle\rangle\langle\langle\tau_i|, \quad \mathbb{P}_2 = \sum_{i=\{1,2,\dots,14\}\setminus\{3,6\}} |\tau_i\rangle\rangle\langle\langle\tau_i|. \quad (\text{B24})$$

The gate-set fidelity in Eq.(B23) shows that gauging its value relies in acquiring an estimate for both $\lambda(\mathbb{P}_1, \Lambda)$ and $\lambda(\mathbb{P}_2, \Lambda)$. The fitting model for each of these quantities follows essentially the same steps as those discussed in Section B 1 a. Indeed, Eqs.(B10) to (B11) carry through, with the exception that the projectors are now either $\mathbb{P}_1$ or $\mathbb{P}_2$, as defined in Eq.(B24). The choice between the two depends on whether we are constructing the fitting model for $\lambda(\mathbb{P}_1, \Lambda)$ or $\lambda(\mathbb{P}_2, \Lambda)$. We note, however, that the convenient choice of input state and POVM elements is no longer the z -computational basis, as for the Clifford group. We can understand this as a consequence of the irrep decomposition of the CNOT-Dihedral group $G(1)$. The states spanning the z -computational basis either belong to the trivial irrep or the 1st non-trivial irrep, with projector $\mathbb{P}_1$. Therefore, this choice of measurement basis fails as good probe to the second non-trivial subspace. This highlights that a *good* choice of POVMs and input state is one that has overlap with all invariant subspaces of the group. For this reason, all simulations employing the CNOT-Dihedral group $G(1)$ were run with the input and POVM elements set to a mixed basis, where the first qubit is measured in the z -computational basis and the second qubit in the x -computational basis, i.e.:

$$\begin{aligned} |\rho\rangle\rangle &= \frac{1}{2} \left(|\tau_0\rangle\rangle + |\tau_1\rangle\rangle + |\tau_6\rangle\rangle + |\tau_{13}\rangle\rangle \right) = |E_{0+}\rangle\rangle, \\ |E_{0-}\rangle\rangle &= \frac{1}{2} \left(|\tau_0\rangle\rangle - |\tau_1\rangle\rangle + |\tau_6\rangle\rangle - |\tau_{13}\rangle\rangle \right), \\ |E_{1+}\rangle\rangle &= \frac{1}{2} \left(|\tau_0\rangle\rangle + |\tau_1\rangle\rangle - |\tau_6\rangle\rangle - |\tau_{13}\rangle\rangle \right), \\ |E_{1-}\rangle\rangle &= \frac{1}{2} \left(|\tau_0\rangle\rangle - |\tau_1\rangle\rangle - |\tau_6\rangle\rangle + |\tau_{13}\rangle\rangle \right). \end{aligned} \quad (\text{B25})$$

2. Interleaved RB

The goal of interleaved RB is to extract information on the average error rate of individual gates [14]. This is achieved through constructing interleaved sequences, where the desired benchmarking gate is intertwined between randomly selected gates. We adopt the same noise modeling convention as in interleaved RB, whereby the noisy

random gates and the noisy target random gate, $\tilde{U}$, are modeled respectively as [14]:

$$\mathcal{R}(\tilde{g}) = \mathcal{R}(\Lambda)\mathcal{R}(g) , \quad \mathcal{R}(\tilde{U}) = \mathcal{R}(U)\mathcal{R}(\Lambda_U) . \quad (\text{B26})$$

Given this noise model and interleaved sequences, the probability distribution, $p(x, \mathbf{g})$, introduced in Eq.(B2) is modified to:

$$\begin{aligned} p(x, \mathbf{g}) &= \langle \langle \tilde{E}_x | \mathcal{R}(\Lambda) \mathcal{R}(g_m) \mathcal{R}(U) \mathcal{R}(\Lambda_U) \mathcal{R}(\Lambda) \mathcal{R}(g_{m-1}) \mathcal{R}(U) \mathcal{R}(\Lambda_U) \mathcal{R}(\Lambda) \dots \mathcal{R}(g_1) | \tilde{\rho} \rangle \rangle \\ &= \langle \langle \tilde{E}_x | \mathcal{R}(\Lambda) \prod_{j=1}^{m-1} \mathcal{R}(\tilde{\Lambda}) \mathcal{R}(g_j) | \tilde{\rho} \rangle \rangle , \end{aligned} \quad (\text{B27})$$

where we defined:

$$\mathcal{R}(\tilde{\Lambda}) := \mathcal{R}(U)\mathcal{R}(\Lambda_U)\mathcal{R}(\Lambda) . \quad (\text{B28})$$

The sequence function model is then given by the expectation value:

$$\begin{aligned} k_A(m) &= \mathbb{E}_{\mathbf{g} \in G^{\times m}} \sum_{x \in \{0,1\}^n} f_A(x, \mathbf{g}) p(x, \mathbf{g}) \\ &= c_{f_A}^{(0)} \mathbb{E}_{\mathbf{g} \in G^{\times m}} \sum_{x \in \{0,1\}^n} \langle \langle E_x | \mathcal{R}_i(g_m) \prod_{j=1}^{m-1} A \mathcal{R}_i(g_j) | \rho \rangle \rangle \langle \langle \tilde{E}_x | \mathcal{R}(\Lambda) \prod_{j=1}^{m-1} \mathcal{R}(\tilde{\Lambda}) \mathcal{R}(g_j) | \tilde{\rho} \rangle \rangle , \end{aligned} \quad (\text{B29})$$

which has a very similar structure to Eq.(B3). This allows us to directly write:

$$k_A(m) = c_{f_A}^{(0)} \sum_{x \in \{0,1\}^n} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\mathbb{P}_{\text{triv}}(A \otimes \mathcal{R}(\tilde{\Lambda})) \mathbb{P}_{\text{triv}} \right)^{m-1} \right] . \quad (\text{B30})$$

To further simplify the fitting model, we need to specify both the choice of gate-set and the probe operator. In accordance with the main text, we examine the fitting model produced by the multi-qubit Clifford group, and the CNOT-dihedral group $G(1)$.

a. Multi-qubit Clifford group

The similarity between Eq.(B30) and Eq.(B6) allows us to immediately write the sequence function for this gate-set as:

$$k_A(m) = c_{f_A}^{(0)} \left(\frac{1}{|\mathbb{P}_1|} \text{Tr}(\mathbb{P}_1 A^T \mathbb{P}_1 \mathcal{R}(\tilde{\Lambda})) \right)^{m-1} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_1|} \langle \langle \mathbb{P}_1 | \Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) | \mathbb{P}_1 \rangle \rangle , \quad (\text{B31})$$

Suppose then that U is one of the Clifford gates in the gate-set. Then choosing the probe operator to be $A = \mathcal{R}(U^*)$, renders the interleaved RB decaying factor as:

$$\begin{aligned} \lambda(U^*, \Lambda_U \Lambda) &= \frac{1}{|\mathbb{P}_1|} \text{Tr}(\mathbb{P}_1 \mathcal{R}^\dagger(U) \mathbb{P}_1 \mathcal{R}(\tilde{\Lambda})) \\ &= \frac{1}{|\mathbb{P}_1|} \text{Tr}(\mathbb{P}_1 \mathcal{R}^\dagger(U) \mathbb{P}_1 \mathcal{R}(U) \mathcal{R}(\Lambda_U) \mathcal{R}(\Lambda)) \\ &= \frac{1}{|\mathbb{P}_1|} \text{Tr}(\mathbb{P}_1 \mathcal{R}(\Lambda_U) \mathcal{R}(\Lambda)) . \end{aligned} \quad (\text{B32})$$

From Eq.(5), the decaying factor $\lambda(U^*, \Lambda_U \circ \Lambda)$ can be directly related to average fidelity $\bar{F}(\Lambda_U \Lambda, \mathcal{I})$. As stated in the main text, having both knowledge on $\bar{F}(\Lambda, \mathcal{I})$ and $\bar{F}(\Lambda_U \Lambda, \mathcal{I})$ is sufficient to bound the value of $\bar{F}(\tilde{U}, U)$. This result arises from the fact that (see Ref. [33] for proof) for any two quantum channels, $\mathcal{A}$ and $\mathcal{B}$, the following inequality

holds true:

$$|\chi_{00}^{(\mathcal{AB})} - \chi_{00}^{(\mathcal{A})} \chi_{00}^{(\mathcal{B})} - (1 - \chi_{00}^{(\mathcal{A})})(1 - \chi_{00}^{(\mathcal{B})})| \leq 2\sqrt{\chi_{00}^{(\mathcal{A})} \chi_{00}^{(\mathcal{B})} (1 - \chi_{00}^{(\mathcal{A})})(1 - \chi_{00}^{(\mathcal{B})})}, \quad (\text{B33})$$

where $\chi_{00}^{(\mathcal{A})} = \frac{(d+1)\bar{F}(\mathcal{A}, \mathcal{I}) - 1}{d}$, and likewise for $\chi_{00}^{(\mathcal{B})}$ and $\chi_{00}^{(\mathcal{AB})}$. By identifying $\mathcal{A} = \Lambda_U$, $\mathcal{B} = \Lambda$, and thus $\mathcal{AB} = \Lambda_U \Lambda$, the inequality can be inverted to obtain an estimate for $\bar{F}(\Lambda_U, \mathcal{I}) = \bar{F}(\tilde{U}, U)$. Note that this approach does not require any assumptions on the specific type of noise affecting the gates, i.e. whether it is a depolarizing channel, a unitary channel, or any other type.

b. CNOT-Dihedral group $G(1)$

Interleaved RB choosing the CNOT-Dihedral group $G(1)$ as gate-set follows similarly to the multi-qubit Clifford group, with the exception that there will be two decaying factors contributing to the average fidelity $\bar{F}(\Lambda_U \Lambda, \mathcal{I})$ (see Eq.(B23)). This follows directly from the decomposition of the group in two non-trivial irreps. Nevertheless, the choice of probe operator remains the same, i.e. $A = \mathcal{R}(U^*)$, so long as U is a gate shared in common by the two gate-sets. Note that the CNOT-Dihedral group $G(1)$ is a subgroup of the two-qubit Clifford group. In the main text, the chosen target gate was a CNOT, which is contained in both groups.

In order to extract each decaying factor separately, two sets of sequence correlation functions, $\{f_A(x_i, \mathbf{g}_i)\}_i$, need to be constructed. The first set is evaluated using only the first non-trivial matrix irrep of the gates, $\mathcal{R}_1(g)$. Instead, the second set is computed using the second non-trivial matrix irrep of the gates, $\mathcal{R}_2(g)$. Note that extracting the matrix irreps can be directly attained from knowledge on the corresponding invariant subspace projector: $\mathcal{R}_i(g) = \mathbb{P}_i \mathcal{R}(g) \mathbb{P}_i$. The last remaining difference with respect to the multi-qubit Clifford group as gate-set is the choice of POVM's and input state, which are now given by Eq.(B25). The desired fidelity $\bar{F}(\tilde{U}, U)$ can then be bounded with Eq.(B33), from knowledge on the estimates of both $\bar{F}(\Lambda_U \Lambda, \mathcal{I})$ (protocol with interleaved sequences) and $\bar{F}(\Lambda, \mathcal{I})$ (protocol with fully random sequences).

3. Simultaneous RB

a. The local Clifford group $\mathbb{C}_1^{\times 2}$

The local Clifford group on two-qubits, $\mathbb{C}_1^{\times 2}$, is another example of a multiplicity-free group. It is formed by taking the direct product of two single-qubit Clifford groups. In the theory of finite representation groups, we can construct the matrix representations of the direct product of two groups by taking the Kronecker product of their irreps [44]. As discussed in Section B1a, the Clifford group has two distinct irreps. Let us then denote $\mathcal{R}_0(g)$ and $\mathcal{R}_1(g)$ as the trivial and non-trivial irreps of single-qubit Clifford group, respectively. The representations $\mathcal{R}_0(g) \otimes \mathcal{R}_0(g)$, $\mathcal{R}_0(g) \otimes \mathcal{R}_1(g)$, $\mathcal{R}_1(g) \otimes \mathcal{R}_0(g)$ and $\mathcal{R}_1(g) \otimes \mathcal{R}_1(g)$ are the resulting direct product representations. We can see that these form irreps of $\mathbb{C}_1^{\times 2}$ by again looking at how these representations transform the normalized Pauli vectors. Note that these are tensor product matrix representations acting on tensor product vectors. As such, each representation acts separately on each single-qubit Pauli, i.e. $(\mathcal{R}_k(g) \otimes \mathcal{R}_l(g))|\sigma_i \otimes \sigma_j\rangle = (\mathcal{R}_k(g)|\sigma_i\rangle) \otimes (\mathcal{R}_l(g)|\sigma_j\rangle)$. Hence, the four representations are irreps with corresponding irreducible subspaces as given in Tab. IV. In simultaneous RB, the

Irreducible Representation	Invariant Subspace	Projector
Trivial irrep	$V_0 = \text{Span}\{ \tau_0\rangle\}$	$\mathbb{P}_0 = \tau_0\rangle\langle\tau_0 $
1st non-trivial irrep	$V_1 = \text{Span}\{ \tau_4\rangle, \tau_5\rangle, \tau_6\rangle\}$	$\mathbb{P}_1 = \sum_{i \in \{4,5,6\}} \tau_i\rangle\langle\tau_i $
2nd non-trivial irrep	$V_2 = \text{Span}\{ \tau_1\rangle, \tau_2\rangle, \tau_3\rangle\}$	$\mathbb{P}_2 = \sum_{i \in \{1,2,3\}} \tau_i\rangle\langle\tau_i $
3rd non-trivial irrep	$V_3 = \text{Span}\{ \tau_7\rangle, \dots, \tau_{15}\rangle\}$	$\mathbb{P}_3 = \sum_{i=7}^{15} \tau_i\rangle\langle\tau_i $

TABLE IV. List of irreps of the $\mathbb{C}_1^{\times 2}$ group. The table also includes the corresponding irreducible subspaces and the projectors to those subspaces. Note that we are adhering to the vector labeling convention introduced in Eq.(A8).

sequence fidelity is a mixture of three decaying parameters, each originating from a depolarizing channel in one of the non-trivial irreducible subspaces [18]. The gate-set shadow protocol allows us to extract each of these parameters separately. To derive the fitting model for each decaying factor, the sequence correlation function $f_A(x, \mathbf{g})$ is evaluated fixing the irrep to one of the non-trivial irreps in Tab. IV. The corresponding fitting model is given by Eq.(B6), with

the trivial projector, $\mathbb{P}_{\text{triv}}$, now related to the invariant subspace of the chosen irrep, i.e. $\mathbb{P}_{\text{triv}} = \frac{1}{|\mathbb{P}_i|} |\mathbb{P}_i\rangle\langle\mathbb{P}_i|$. Thus, this leads to a decaying model similar to Eq.(B11). By fixing the probe operator A as the projector onto the invariant subspace of the chosen irrep, we arrive at one of the fitting models in Tab. II. Proceeding in this way for all non-trivial irreps, we can generate each of the fitting models in Tab. II. Each model is of the form:

$$k_A(m) = c_i \left(\frac{\text{Tr}(\mathbb{P}_i \Lambda)}{|\mathbb{P}_i|} \right)^{m-1}. \quad (\text{B34})$$

The constant c_i is a SPAM dependent parameter. From Eq.(B11), we can immediately conclude that c_i is given by:

$$c_i = c_{f_{A_i}}^{(0)} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_i|} \langle\langle \mathbb{P}_i | \Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) | \mathbb{P}_i \rangle\rangle, \quad (\text{B35})$$

with $\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x)$ defined as in section B. The set of normalization constants $\{c_{f_{A_i}}^{(0)}\}$ is determined by the limit of no SPAM and gate errors. A priori, each irrep can give rise to a different normalization constant. In this case, and using the input state and POVM elements defined in Eq.(B14), we have:

$$\begin{aligned} A = \mathbb{P}_1 : k_A(1) &= c_{f_{\mathbb{P}_1}}^{(0)} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_1|} \langle\langle \mathbb{P}_1 | \Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) | \mathbb{P}_1 \rangle\rangle \\ &= \frac{c_{f_{\mathbb{P}_1}}^{(0)}}{3} \sum_{i \in \{4,5,6\}} \langle\langle \tau_i | \rho \rangle\rangle^2 \sum_{x \in \{0,1\}^2} \sum_{j \in \{4,5,6\}} \langle\langle E_x | \tau_j \rangle\rangle^2 \\ &= \frac{c_{f_{\mathbb{P}_1}}^{(0)}}{12} \implies c_{f_{\mathbb{P}_1}}^{(0)} = 12. \end{aligned} \quad (\text{B36})$$

$$\begin{aligned} A = \mathbb{P}_3 : k_A(1) &= c_{f_{\mathbb{P}_3}}^{(0)} \sum_{x \in \{0,1\}^2} \frac{1}{|\mathbb{P}_3|} \langle\langle \mathbb{P}_3 | \Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) | \mathbb{P}_3 \rangle\rangle \\ &= \frac{c_{f_{\mathbb{P}_3}}^{(0)}}{9} \sum_{i=7}^{15} \langle\langle \tau_i | \rho \rangle\rangle^2 \sum_{x \in \{0,1\}^2} \sum_{j=7}^{15} \langle\langle E_x | \tau_j \rangle\rangle^2 \\ &= \frac{c_{f_{\mathbb{P}_3}}^{(0)}}{36} \implies c_{f_{\mathbb{P}_3}}^{(0)} = 36. \end{aligned} \quad (\text{B37})$$

Following similar arguments, $c_{f_{\mathbb{P}_2}}^{(0)} = 12$.

b. The group $\mathbb{C}_1 \times \mathcal{I}$

The group $\mathbb{C}_1 \times \mathcal{I}$ is a non-multiplicity-free group, and is a subgroup of the local Clifford group $\mathbb{C}_1^{\times 2}$. The non-multiplicity arises because there are four copies for each of the two distinct irreps of the $\mathbb{C}_1 \times \mathcal{I}$ group. To motivate this statement, let us consider only the subset of operations of the form $\mathcal{R}(g) \otimes \mathbb{1} \in \mathbb{C}_1^{\times 2}$, and ask how these transform the normalized Pauli vectors in each invariant subspace of the local Clifford group (see Tab. IV for definition of the subspaces). We start with the trivial subspace, spanned by the vector $|\sigma_0 \otimes \sigma_0\rangle\rangle$, and note that it is again mapped to itself by the action of all $\mathcal{R}(g) \otimes \mathbb{1}$ operations. The second non-trivial invariant subspace, however, is partitioned into three vector spaces. This can be understood by noting that the vectors $|\sigma_0 \otimes \sigma_x\rangle\rangle$, $|\sigma_0 \otimes \sigma_y\rangle\rangle$ and $|\sigma_0 \otimes \sigma_z\rangle\rangle$ are all mapped to themselves by the operations $\mathcal{R}(g) \otimes \mathbb{1}$. Therefore, the group $\mathbb{C}_1 \times \mathcal{I}$ gives rise to four invariant vector spaces for the trivial irrep. Proceeding in the same fashion for the remaining irreps of the local Clifford group, it is possible to conclude that the vector subspace of the first non-trivial irrep prevails as an invariant subspace for the action of the operations $\mathcal{R}(g) \otimes \mathbb{1}$. The same is no longer true for the third invariant subspace. Here, the action of the operations $\mathcal{R}(g) \otimes \mathbb{1}$ partitions the subspace into three smaller invariant subspaces: $\{|\sigma_i \otimes \sigma_x\rangle\rangle\}_{i \in \{x,y,z\}}$, $\{|\sigma_i \otimes \sigma_y\rangle\rangle\}_{i \in \{x,y,z\}}$ and $\{|\sigma_i \otimes \sigma_z\rangle\rangle\}_{i \in \{x,y,z\}}$. These vector spaces, along with subspace of the first non-trivial irrep, all correspond to the same non-trivial irrep: they all describe how the action of the group permutes (up to a phase) the basis vectors for

the first qubit state. Thus, in $\mathbb{C}_1 \times \mathcal{I}$ the non-trivial irrep occurs with multiplicity four. Note that another approach to ascertaining the multiplicity of the irreps is to directly evaluate Eq.(A15) numerically. For the task of estimating the fidelity in Eq.(9), the convenient choice of probe operator is to set $A = \mathbb{P}_1$, with the projector $\mathbb{P}_1$ defined as in Tab. IV. This requires us to build sequence correlation functions, where the gate operations are represented by the non-trivial irrep. Given the freedom of basis choice, we take the non-trivial irrep to be expressed in the same basis as the probe operator. This corresponds to fixing the irrep to one specific copy in $f_A(x, \mathbf{g})$. To obtain the corresponding fitting model, we need to evaluate the average sequence function $k_A(m)$. All steps in Section B are generic, and thus, carry through. The choice of gate-set, however, dictates the basis decomposition of the trivial projector, $\mathbb{P}_{\text{triv}}$, which now contains mixing of the different irrep copies. To simplify notation, let us re-label the basis vectors in Eq.(A8) as follows:

Non-trivial irrep, copy 1:

$$|\sigma_x \otimes \sigma_0\rangle \equiv |\tau_1^{(1)}\rangle, |\sigma_y \otimes \sigma_0\rangle \equiv |\tau_2^{(1)}\rangle, |\sigma_z \otimes \sigma_0\rangle \equiv |\tau_3^{(1)}\rangle,$$

Non-trivial irrep, copy 2:

$$|\sigma_x \otimes \sigma_x\rangle \equiv |\tau_1^{(2)}\rangle, |\sigma_y \otimes \sigma_x\rangle \equiv |\tau_2^{(2)}\rangle, |\sigma_z \otimes \sigma_x\rangle \equiv |\tau_3^{(2)}\rangle,$$

Non-trivial irrep, copy 3:

$$|\sigma_x \otimes \sigma_y\rangle \equiv |\tau_1^{(3)}\rangle, |\sigma_y \otimes \sigma_y\rangle \equiv |\tau_2^{(3)}\rangle, |\sigma_z \otimes \sigma_y\rangle \equiv |\tau_3^{(3)}\rangle.$$

Non-trivial irrep, copy 4:

$$|\sigma_x \otimes \sigma_z\rangle \equiv |\tau_1^{(3)}\rangle, |\sigma_y \otimes \sigma_z\rangle \equiv |\tau_2^{(3)}\rangle, |\sigma_z \otimes \sigma_z\rangle \equiv |\tau_3^{(3)}\rangle.$$

(B38)

In Eq.(B38), the basis vectors spanning the trivial irrep are not included. This is justified, since the trivial projector will not have any contribution from this irrep. For this reason, we neglect the trivial irrep and its copies for the remaining of this section.

The vector basis of $\mathbb{P}_{\text{triv}}$ in Eq.(B6) is now given by $\{|\tau_l^{(1)}\rangle \otimes |\tau_l^{(j)}\rangle\}$. Recall from Eq.(B5) that the trivial projector was defined by averaging over representations of the form $\mathcal{R}_i(g) \otimes \mathcal{R}_i(g)$, where $\mathcal{R}_i(g)$ is the i th non-trivial irrep. Since the group is now non-multiplicity-free, we define $\mathcal{R}_1^{(i)}(g)$ as the non-trivial irrep matrix in the i th copy subspace. Furthermore, because we fix the subspace in which the sequence correlation functions are evaluated, we need to search for a set of independent vectors that are invariant under the matrix operations of the form $\mathcal{R}_1^{(1)}(g) \otimes \mathcal{R}_1^{(i)}(g)$. These vectors will provide a basis for the trivial projector $\mathbb{P}_{\text{triv}}$, namely:

$$\mathbb{P}_{\text{triv}} = \frac{1}{|\mathbb{P}_1|} \sum_{j=1}^{n_1} \vec{v}_{\text{triv}}^{(j)} (\vec{v}_{\text{triv}}^{(j)})^\dagger = \frac{1}{|\mathbb{P}_1|} \sum_{l,k=1}^{|\mathbb{P}_1|} |\mathbb{P}_l^{(1,j)}\rangle \langle \mathbb{P}_k^{(1,j)}|,$$

$$\text{with } \vec{v}_{\text{triv}}^{(j)} = \frac{1}{\sqrt{|\mathbb{P}_1|}} \sum_{l=1}^{|\mathbb{P}_1|} |\mathbb{P}_l^{(1,j)}\rangle = \frac{1}{\sqrt{|\mathbb{P}_1|}} \sum_{l=1}^{|\mathbb{P}_1|} \text{vec}(|\tau_l^{(1)}\rangle \langle \tau_l^{(j)}|) = \frac{1}{\sqrt{|\mathbb{P}_1|}} \sum_{l=1}^{|\mathbb{P}_1|} |\tau_l^{(1)}\rangle \otimes |\tau_l^{(j)}\rangle.$$

(B39)

Hence, the fitting model in Eq.(B6) simplifies to:

$$\begin{aligned}
k_A(m) &= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\frac{1}{|\mathbb{P}_1|^2} \sum_{k,k'} \sum_{l,l',n,n'} \langle \mathbb{P}_{l'}^{(1,k)} | (A \otimes \mathcal{R}(\Lambda)) | \mathbb{P}_n^{(1,k')} \rangle \langle \mathbb{P}_l^{(1,k)} \rangle \langle \mathbb{P}_{n'}^{(1,k')} | \right)^{m-1} \right] \\
&= c_{f_A}^{(0)} \sum_{x \in \{0,1\}^2} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\frac{1}{|\mathbb{P}_1|^2} \sum_{k,k'} \sum_{l,l',n,n'} \langle \mathbb{P}_{l'}^{(1,k)} | \mathcal{R}(\Lambda) \mathbb{P}_n^{(1,k')} A^T \rangle \langle \mathbb{P}_l^{(1,k)} \rangle \langle \mathbb{P}_{n'}^{(1,k')} | \right)^{m-1} \right] \\
&= \frac{c_{f_A}^{(0)}}{|\mathbb{P}_1|} \sum_{x \in \{0,1\}^2} \text{Tr} \left[\Omega(\rho, \tilde{\rho}, E_x, \tilde{E}_x) \left(\lambda_{1,1} \sum_{l,l'} |\mathbb{P}_l^{(1,1)} \rangle \langle \mathbb{P}_{l'}^{(1,1)}| \right)^{m-1} \right] \\
&= \frac{c_{f_A}^{(0)}}{|\mathbb{P}_1|} \sum_{x \in \{0,1\}^2} \text{Tr} \left[\lambda_{1,1}^{m-1} |\rho \otimes \tilde{\rho}\rangle \langle \langle E_x \otimes \Lambda^*(\tilde{E}_x) | \sum_{l,l'} |\mathbb{P}_l^{(1,1)} \rangle \langle \mathbb{P}_{n'}^{(1,1)}| \right] \\
&= \frac{c_{f_A}^{(0)}}{|\mathbb{P}_1|} \sum_{x \in \{0,1\}^2} \lambda_{1,1}^{m-1} \sum_{l,l'} \langle \langle E_x \otimes \Lambda^*(\tilde{E}_x) | \mathbb{P}_l^{(1,1)} \rangle \rangle \langle \langle \mathbb{P}_{l'}^{(1,1)} | \rho \otimes \tilde{\rho} \rangle \rangle \\
&= \frac{c_{f_A}^{(0)}}{|\mathbb{P}_1|} \sum_{x \in \{0,1\}^2} \lambda_{1,1}^{m-1} \sum_{l,l'} \langle \langle \tau_{l'}^{(1)} | \rho \rangle \rangle \langle \langle E_x | \tau_l^{(1)} \rangle \rangle \langle \langle \tau_{l'}^{(1)} | \tilde{\rho} \rangle \rangle \langle \langle \tilde{E}_x | \mathcal{R}(\Lambda) | \tau_l^{(1)} \rangle \rangle ,
\end{aligned} \tag{B40}$$

where we have used the choice of probe operator $A = \mathbb{P}_1$ and applied the vectorization identity in Eq.(B12). Note that $\lambda_{k,k'}$ is defined as:

$$\begin{aligned}
\lambda_{k,k'} &= \frac{1}{|\mathbb{P}_1|} \sum_{l',n} \langle \langle \mathbb{P}_{l'}^{(1,k)} | \mathcal{R}(\Lambda) \mathbb{P}_n^{(1,k')} A^T \rangle \rangle \\
&= \frac{1}{|\mathbb{P}_1|} \sum_{l',n} \text{Tr} (\mathbb{P}_{l'}^{(k,1)} \mathcal{R}(\Lambda) \mathbb{P}_n^{(1,k')} A^T) \\
&= \frac{1}{|\mathbb{P}_1|} \sum_{l',n} \langle \langle \tau_{l'}^{(1)} | \mathcal{R}(\Lambda) | \tau_n^{(1)} \rangle \rangle \langle \langle \tau_n^{(k')} | A^T | \tau_{l'}^{(k)} \rangle \rangle \\
&= \frac{1}{|\mathbb{P}_1|} \sum_{j=1}^3 \langle \langle \tau_j^{(1)} | \mathcal{R}(\Lambda) | \tau_j^{(1)} \rangle \rangle \delta_{k',1} \delta_{k,1} .
\end{aligned} \tag{B41}$$

In the main text, we have defined $\lambda_1 \equiv \lambda_{1,1}$. The POVM elements and input state, defined in Eq.(10), are given by:

$$\begin{aligned}
|\rho\rangle\rangle &= |\sigma_0 \otimes \sigma_0\rangle\rangle + |\tau_3^{(1)}\rangle\rangle , \\
|E_0\rangle\rangle &= 2|\rho\rangle\rangle , \\
|E_1\rangle\rangle &= |\sigma_0 \otimes \sigma_0\rangle\rangle - |\tau_3^{(1)}\rangle\rangle ,
\end{aligned} \tag{B42}$$

and can be used to simplify the sequence function, $k_A(m)$, further. Applying the definitions in Eq.(B42) to Eq.(B40) leads to the sequence function:

$$\begin{aligned}
k_A(m) &= \frac{c_{f_A}^{(0)}}{2|\mathbb{P}_1|} \sum_{x \in \{0,1\}^2} \lambda_{1,1}^{m-1} \sum_{l,l'} \delta_{l',3} \delta_{l,3} (-1)^x \langle \langle \tau_{l'}^{(1)} | \tilde{\rho} \rangle \rangle \langle \langle \tilde{E}_x | \mathcal{R}(\Lambda) | \tau_l^{(1)} \rangle \rangle \\
&= \frac{c_{f_A}^{(0)}}{2|\mathbb{P}_1|} \lambda_{1,1}^{m-1} \langle \langle \tau_3^{(1)} | \tilde{\rho} \rangle \rangle \langle \langle (\tilde{E}_0 - \tilde{E}_1) | \mathcal{R}(\Lambda) | \tau_3^{(1)} \rangle \rangle .
\end{aligned} \tag{B43}$$

Thus, SPAM errors can affect the fitting procedure, by reducing the amplitude of the multiplicative pre-factor. The fitting model, however, remains a single decay model even in the presence of SPAM. In the absence of SPAM errors,

the model reduces to:

$$k_A(m) = \frac{c_{f_A}^{(0)}}{2|\mathbb{P}_1|} \lambda_{1,1}^{m-1} \langle \langle \tau_3^{(1)} | \mathcal{R}(\Lambda) | \tau_3^{(1)} \rangle \rangle , \quad (\text{B44})$$

which, in the limit of perfect gate operations, allows to determine the normalization constant as $c_{f_A}^{(0)} = 2|\mathbb{P}_1|$.

4. Leakage RB

The aim of leakage RB is two-fold: to characterize the leakage and seepage rates and to estimate the average gate fidelity restricted to the computational subspace. In this work, we focused only on the first goal. Recall that the leakage and seepage rates are defined as [27]:

$$\begin{aligned} L &= \frac{1}{d_1} \langle \langle \mathbb{1}_2 | \Lambda | \mathbb{1}_1 \rangle \rangle , \\ S &= \frac{1}{d_2} \langle \langle \mathbb{1}_1 | \Lambda | \mathbb{1}_2 \rangle \rangle . \end{aligned} \quad (\text{B45})$$

In this case, $d_1 = d_2 = 2$. Since we want to estimate both L and S , this anticipates that a useful choice of probe operator is to take a linear combination of the computational and leakage subspace projectors, $|\mathbb{1}_1\rangle\rangle\langle\langle\mathbb{1}_1|$ and $|\mathbb{1}_2\rangle\rangle\langle\langle\mathbb{1}_2|$, respectively. The gate-set group here is given by the generators in Eq.(16) [27], and contains 16 distinct elements, given by:

$$\begin{aligned} G = \{ & \mathbb{1} \oplus \mathbb{1}, \mathbb{1} \oplus (-\mathbb{1}), X \oplus Z, X \oplus (-Z), Z \oplus H, Z \oplus (-H), \\ & X \oplus X, X \oplus (-X), \mathbb{1} \oplus iY, \mathbb{1} \oplus (-iY), iY \oplus HZ, \\ & iY \oplus (-HZ), iY \oplus HX, iY \oplus (-HX), Z \oplus iHY, Z \oplus (-iHY) \} . \end{aligned} \quad (\text{B46})$$

This group is a non-multiplicity-free group. Additionally, while the group elements induce separate transformations on the computational and leakage subspace, G cannot be written as a direct sum group [27]. This means that the irrep decomposition of the group does not align with the partition of the vector space into a computational and leakage subspaces: the irreps' invariant subspaces can mix both the computational and leakage subspace. Indeed, this statement is already true for the trivial irrep. This irrep has two copies, each spanned by a one-dimensional vector space containing either the basis vector $|\sigma_0 \otimes \sigma_0\rangle\rangle$ or $|\sigma_z \otimes \sigma_0\rangle\rangle$. Note that we can equivalently write $\sigma_0 \otimes \sigma_0 = \sigma_0 \oplus \sigma_0$ and $\sigma_z \otimes \sigma_0 = \sigma_0 \oplus (-\sigma_0)$, from which follows that the group elements leave the states $|\sigma_0 \otimes \sigma_0\rangle\rangle$ and $|\sigma_z \otimes \sigma_0\rangle\rangle$ unchanged. Another crucial observation is that the states $|\mathbb{1}_1\rangle\rangle$ and $|\mathbb{1}_2\rangle\rangle$ admit a full decomposition in terms of the PTM states spanning the trivial irrep:

$$|\mathbb{1}_1\rangle\rangle = |\sigma_0 \otimes \sigma_0\rangle\rangle + |\sigma_z \otimes \sigma_0\rangle\rangle , \quad |\mathbb{1}_2\rangle\rangle = |\sigma_0 \otimes \sigma_0\rangle\rangle - |\sigma_z \otimes \sigma_0\rangle\rangle \quad (\text{B47})$$

This guides our choice of sequence correlation function to be one in which the irreducible representation is fixed to the trivial irrep. However, unlike in Section B3b, we now take the trivial irrep as given by the block diagonal portion containing the two copies of the trivial irrep. Alternatively, we can view this sequence correlation function as the sum of two *elementary* sequence correlation functions, where for each the trivial irrep is fixed to one of the irrep's copy. We define as in the main text the projectors:

$$\begin{aligned} \mathbb{P}_0^{(1)} &\equiv \mathbb{P}_0^{(1,1)} := |\sigma_0 \otimes \sigma_0\rangle\rangle\langle\langle\sigma_0 \otimes \sigma_0| , \\ \mathbb{P}_0^{(1,2)} &:= |\sigma_0 \otimes \sigma_0\rangle\rangle\langle\langle\sigma_z \otimes \sigma_0| , \\ \mathbb{P}_0^{(2,1)} &:= |\sigma_z \otimes \sigma_0\rangle\rangle\langle\langle\sigma_0 \otimes \sigma_0| , \\ \mathbb{P}_0^{(2)} &\equiv \mathbb{P}_0^{(2,2)} := |\sigma_z \otimes \sigma_0\rangle\rangle\langle\langle\sigma_z \otimes \sigma_0| , \end{aligned} \quad (\text{B48})$$

and consider two choices for the probe operator: $A = \mathbb{P}_0^{(1)} + \mathbb{P}_0^{(2)} \equiv \mathbb{P}_0$ or $A = \mathbb{P}_0^{(2)}$. The first choice leads to the sequence correlation function:

$$\begin{aligned} f_{\mathbb{P}_0}(x, \mathbf{g}) &= c_{f_A}^{(0)} \langle \langle E_x | \mathcal{R}_i(g_m) \prod_{j=1}^{m-1} \mathbb{P}_0 \mathcal{R}_i(g_j) | \rho \rangle \rangle = c_{f_A}^{(0)} \langle \langle E_x | \mathbb{P}_0^{(1)} | \rho \rangle \rangle + c_{f_A}^{(0)} \langle \langle E_x | \mathbb{P}_0^{(2)} | \rho \rangle \rangle \\ &= c_{f_A}^{(0)} \delta_{E_x=1_1}, \end{aligned} \quad (\text{B49})$$

where we have used the ideal set of POVM elements and input state, defined as:

$$|E_x\rangle\rangle = \{|E_0\rangle\rangle, |E_1\rangle\rangle\} = \{|\mathbb{1}_1\rangle\rangle, |\mathbb{1}_2\rangle\rangle\}, \quad |\rho\rangle\rangle = \frac{1}{2}|\mathbb{1}_1\rangle\rangle. \quad (\text{B50})$$

Note that since we are taking the trivial irrep as the block diagonal portion containing both copies of the trivial irrep, its matrix representation is equivalent to the matrix representation of the projector $\mathbb{P}_0$. The combined action of the sequence of gates and interleaved probe is then reduced to the action of the single projector in Eq.(B49). The second choice of probe operator is equivalent to only retaining one of the terms in Eq.(B49):

$$f_{\mathbb{P}_0^{(2)}}(x, \mathbf{g}) = c_{f_A}^{(0)} \langle \langle E_x | \mathbb{P}_0^{(2)} | \rho \rangle \rangle = \frac{c_{f_A}^{(0)}}{2} (\delta_{E_x=1_1} - \delta_{E_x=1_2}). \quad (\text{B51})$$

The simplified sequence correlation function in Eq.(B49) leads to the sequence function:

$$\begin{aligned} k_{\mathbb{P}_0}(m) &= c_{f_A}^{(0)} \mathbb{E}_{\mathbf{g} \in G^{\times m}} \langle \langle \tilde{E}_0 | \prod_{j=1}^m \mathcal{R}(\Lambda) \mathcal{R}(g_j) | \tilde{\rho} \rangle \rangle = \\ &= c_{f_A}^{(0)} \mathbb{E}_{\mathbf{g} \in G^{\times m}} \langle \langle \Lambda^*(\tilde{E}_0) | D(g_m) \prod_{j=1}^{m-1} (D^\dagger(g_j) \mathcal{R}(\Lambda) D(g_j)) | \tilde{\rho} \rangle \rangle \\ &= c_{f_A}^{(0)} \langle \langle \Lambda^*(\tilde{E}_0) | \left(\mathbb{E}_{g_m \in G} D(g_m) \right) \left(\mathbb{E}_{g \in G} D^\dagger(g) \mathcal{R}(\Lambda) D(g) \right)^{m-1} | \tilde{\rho} \rangle \rangle \\ &= c_{f_A}^{(0)} \langle \langle \Lambda^*(\tilde{E}_0) | \mathbb{P}_0 \left(\mathbb{E}_{g \in G} D^\dagger(g) \mathcal{R}(\Lambda) D(g) \right)^{m-1} | \tilde{\rho} \rangle \rangle \\ &= c_{f_A}^{(0)} \langle \langle \Lambda^*(\tilde{E}_0) | \mathbb{P}_0 \left(\mathbb{P}_0 \mathbb{E}_{g \in G} D^\dagger(g) \mathcal{R}(\Lambda) D(g) \right)^{m-1} | \tilde{\rho} \rangle \rangle \\ &= c_{f_A}^{(0)} \langle \langle \Lambda^*(\tilde{E}_0) | \left[\mathbb{P}_0^{(1,1)} + \langle \langle \sigma_z \otimes \sigma_0 | \mathcal{R}(\Lambda) | \sigma_0 \otimes \sigma_0 \rangle \rangle \mathbb{P}_0^{(2,1)} + \langle \langle \sigma_z \otimes \sigma_0 | \mathcal{R}(\Lambda) | \sigma_z \otimes \sigma_0 \rangle \rangle \mathbb{P}_0^{(2,2)} \right]^{m-1} | \tilde{\rho} \rangle \rangle, \end{aligned} \quad (\text{B52})$$

where we have re-written the gate sequence in terms of the new random gates, defined recursively as:

$$\begin{aligned} D(g_1) &= \mathcal{R}(g_1), \\ D(g_{i+1}) &= \mathcal{R}(g_{i+1}) D(g_i), \end{aligned} \quad (\text{B53})$$

and used the group averages in Eq.(A16) and Eq.(A18) to obtain the trivial projector $\mathbb{P}_0$ and the average group twirling channel, respectively. Note that the re-parametrization of the gates in Eq.(B53) is akin to the procedure followed in standard RB [12]. Focusing on the operator product in Eq.(B52):

$$\begin{aligned} &\left[\mathbb{P}_0^{(1,1)} + \langle \langle \sigma_z \otimes \sigma_0 | \mathcal{R}(\Lambda) | \sigma_0 \otimes \sigma_0 \rangle \rangle \mathbb{P}_0^{(2,1)} + \langle \langle \sigma_z \otimes \sigma_0 | \mathcal{R}(\Lambda) | \sigma_z \otimes \sigma_0 \rangle \rangle \mathbb{P}_0^{(2,2)} \right]^{m-1} = \\ &= \mathbb{P}_0^{(1,1)} + (\mathcal{R}_{6,6}(\Lambda))^{m-1} \mathbb{P}_0^{(2,2)} + \mathcal{R}_{6,0}(\Lambda) \left(\sum_{n=0}^{m-2} (\mathcal{R}_{6,6}(\Lambda))^n \right) \mathbb{P}_0^{(2,1)} = \\ &= \mathbb{P}_0^{(1,1)} + (\mathcal{R}_{6,6}(\Lambda))^{m-1} \mathbb{P}_0^{(2,2)} + \mathcal{R}_{6,0}(\Lambda) \left(\frac{1 - \mathcal{R}_{6,6}^{m-1}(\Lambda)}{1 - \mathcal{R}_{6,6}(\Lambda)} \right) \mathbb{P}_0^{(2,1)}, \end{aligned} \quad (\text{B54})$$

where we employed the vector basis labeling defined in Eq.(A8) and, thus, changed to a more compact notation, with $\mathcal{R}_{6,6}(\Lambda) = \langle \langle \tau_6 | \mathcal{R}(\Lambda) | \tau_6 \rangle \rangle$ and $\mathcal{R}_{6,0}(\Lambda) = \langle \langle \tau_6 | \mathcal{R}(\Lambda) | \tau_0 \rangle \rangle$. Note that in Eq.(B54), the term multiplying $\mathcal{R}_{6,0}(\Lambda)$ is a geometric series, with constant ration $|r| < 1$, whenever gate errors are present. In the limit of no gate errors, the

geometric series converges to $m - 1$, and Eq.(B54) simplifies to the projector $\mathbb{P}_0$. This limit is instructive to fix the value of the normalization constant $c_{f_A}^{(0)}$: inserting $\mathbb{P}_0$ in Eq.(B52) renders $c_{f_A}^{(0)} = 1$. In the presence of noisy gates, the operator product is given by Eq.(B54), which can then be replaced in Eq.(B52) to produce:

$$\begin{aligned} k_{\mathbb{P}_0}(m) &= b_0 + a_0 \lambda^{m-1}, \text{ with:} \\ b_0 &= \langle \langle \tau_0 | \tilde{\rho} \rangle \rangle \left(\langle \langle \tilde{E}_0 | \mathcal{R}(\Lambda) | \tau_0 \rangle \rangle + \frac{\mathcal{R}_{6,0}(\Lambda)}{1 - \mathcal{R}_{6,6}(\Lambda)} \langle \langle \tilde{E}_0 | \mathcal{R}(\Lambda) | \tau_6 \rangle \rangle \right), \\ a_0 &= \langle \langle \tilde{E}_0 | \mathcal{R}(\Lambda) | \tau_6 \rangle \rangle \left(\langle \langle \tau_6 | \tilde{\rho} \rangle \rangle - \frac{\mathcal{R}_{6,0}(\Lambda)}{1 - \mathcal{R}_{6,6}(\Lambda)} \langle \langle \tau_0 | \tilde{\rho} \rangle \rangle \right), \\ \lambda &= \mathcal{R}_{6,6}(\Lambda). \end{aligned} \quad (\text{B55})$$

To clarify the relationship between the constants a_0 and b_0 and the leakage and seepage rates, we need information on SPAM noise. This underscores the primary challenge in estimating these parameters: in the case of leakage RB, they appear as multiplicative constants rather than as the exponent that sets the decay rate of the sequence function. Consequently, their estimates are necessarily affected by the presence of SPAM noise. Assuming that no SPAM noise is present, Eq.(B55) reduces to:

$$k_{\mathbb{P}_0}(m) = \frac{S}{S + L} + \lambda^m \frac{L}{L + S}, \quad (\text{B56})$$

where we have used the following additional equalities between the noise matrix elements and the leakage and seepage rates:

$$\begin{aligned} 1 - L &= \frac{1}{2} \langle \langle \mathbb{1}_1 | \mathcal{R}(\Lambda) | \mathbb{1}_1 \rangle \rangle, \\ 1 - L - S &= \mathcal{R}_{6,6}(\Lambda) = \lambda, \\ S - L &= \mathcal{R}_{6,0}(\Lambda). \end{aligned} \quad (\text{B57})$$

For the second choice of probe operator, $A = \mathbb{P}_0^{(2)}$, Eq.(B51) allows the sequence function to be expressed as:

$$k_{\mathbb{P}_0^{(2)}}(m) = k_{\mathbb{P}_0}(m) - \mathbb{E}_{\mathbf{g} \in G^{\times m}} \langle \langle \tilde{E}_1 | \prod_{j=1}^m \mathcal{R}(\Lambda) \mathcal{R}(g_j) | \tilde{\rho} \rangle \rangle. \quad (\text{B58})$$

The second term in this equation has essentially the same form as Eq.(B55), with the replacement $\tilde{E}_0 \rightarrow \tilde{E}_1$. Following then a similar procedure, in the limit of no SPAM errors, we obtain:

$$k_{\mathbb{P}_0^{(2)}}(m) = \frac{(S - L)}{S + L} + \frac{2L}{L + S} \lambda^m. \quad (\text{B59})$$

Since $S - L = \mathcal{R}_{6,0}(\Lambda)$, for all unital gate noise channels, Eq.(B59) simplifies to a simpler fitting model.

Appendix C: Confidence Intervals

In general terms, bootstrap methods entail resampling from the original data, with replacement, to generate replicate datasets. For each replicate dataset, we can construct an estimate of the target parameter, in this case $\hat{\lambda}_i$. In the context of our problem, a full bootstrap loop consists of the following iterative steps. First, creating replicate datasets by resampling, with replacement, from the set of sequence correlation functions $\{f(x_i, \mathbf{g}_i)\}_{i=1}^K$, with fixed sequence length m . This step is performed for all values of m probed during the simulated experiment. The attained replicate correlation function datasets are then used to obtain the average sequence function $\hat{k}_A(m)$, which, following the usual procedure, is fitted to a decay model to generate an estimate for $\hat{\lambda}_i$. This entire process is repeated B times, culminating in the creation of B bootstrap estimates for λ_i . We fixed the number of bootstrap samples to be $B = 10000$, since it is typically recommended to use a higher value of resampling for the estimation confidence intervals [36]. We obtain estimates for the confidence intervals using two non-parametric bootstrap methods: the percentile method and the BCa (bias-corrected and accelerated) method [35, 36]. Additionally, we also employed the bootstrap procedure to estimate the sample variance. The sample variance was then used and to construct the

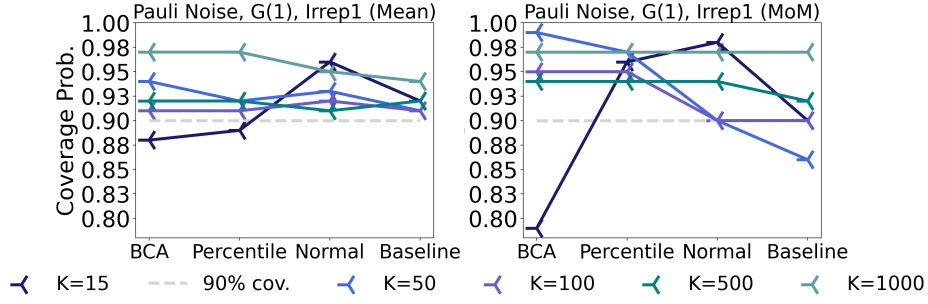


FIG. 10. Coverage probability for different confidence intervals and different sample sizes. The results are obtained by simulating the gate-set shadow protocol 100 times, and recomputing the confidence intervals, using each of the listed methods: the BCa method, the percentile method, taking the normal distribution assumption, and a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm. In this case, the protocol uses the CNOT-dehedral group, $G = G(1)$, as gate-set and gate errors are simulated with a Pauli noise channel. Recall that this group has two decay factors λ_i . The present case shows the coverage probability for λ_1 .

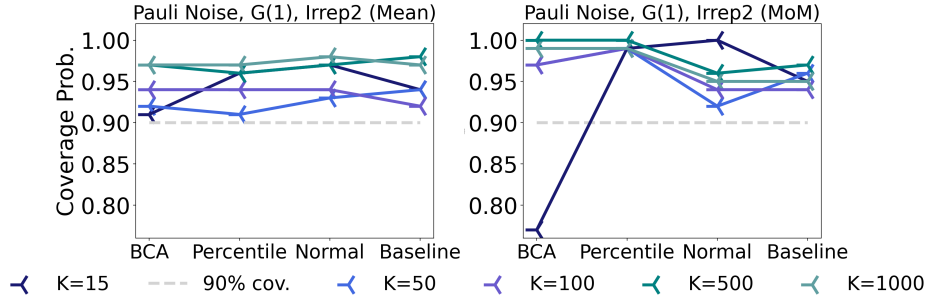


FIG. 11. Coverage probability for different confidence intervals and different sample sizes. The results are obtained by simulating the gate-set shadow protocol 100 times, and recomputing the confidence intervals, using each of the listed methods: the BCa method, the percentile method, taking the normal distribution assumption, and a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm. In this case, the protocol uses the CNOT-dehedral group, $G = G(1)$, as gate-set and gate errors are simulated with a Pauli noise channel. Recall that this group has two decay factors λ_i . The present case shows the coverage probability for λ_2 .

confidence interval resulting from assuming the normal distribution. Lastly, we also obtained the confidence interval resulting from bounding the target estimate by a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm. In this section, we offer additional examples of the coverage probability as a function the sample size for the Clifford group, C_2 , or the CNOT-Dihedral group, $G = G(1)$. In Figs.(10)-(14), we show the coverage probability generated by each confidence interval, across different values of K . The confidence intervals in Figs.(10)-(14) correspond to different methods of obtaining a confidence region from a finite date-set. Across all Figs.(10)-(14), if the sample size is big enough, then all approximate confidence interval are conservative, i.e. the actual coverage probability is above the nominal value (here set as 90% coverage probability). If we focus on the empirical mean and the Clifford group as gate-set, in general we would require a very large value of K to attain a confidence interval with at least 90% coverage probability, based on the standard distribution. This situation generally improves, if we consider the smaller gate-set provided by the CNOT-dihedral group. For all Figs.(10)-(14), the MoM estimator provides more conservative confidence intervals, even for smaller sample sizes. The BCa method gives consistently more conservative confidence intervals, across all sample sizes, and for the error channels explored in this work.

Appendix D: Numerical simulation details

In this work, all numerical simulations pertaining the experimental stage of the gate-set shadow protocol were obtained using the PennyLane Python package [49].

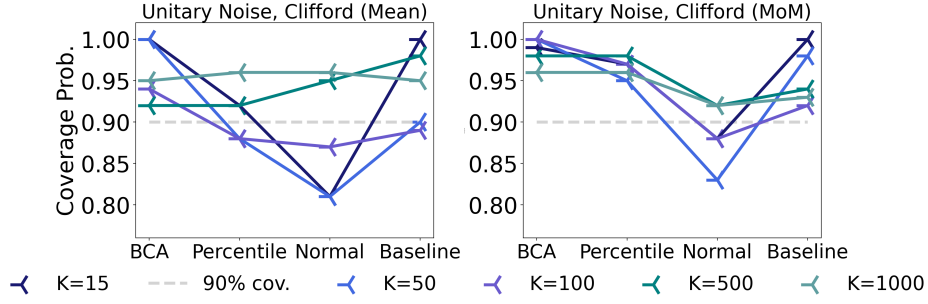


FIG. 12. Coverage probability for different confidence intervals and different sample sizes. The results are obtained by simulating the gate-set shadow protocol 100 times, and recomputing the confidence intervals, using each of the listed methods: the BCa method, the percentile method, taking the normal distribution assumption, and a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm. In this case, the protocol uses the two-qubit Clifford group as gate-set and gate errors are simulated with a unitary channel, namely the two-qubit zz gate $R_{zz}(\theta)$, with $\theta = 0.1$.

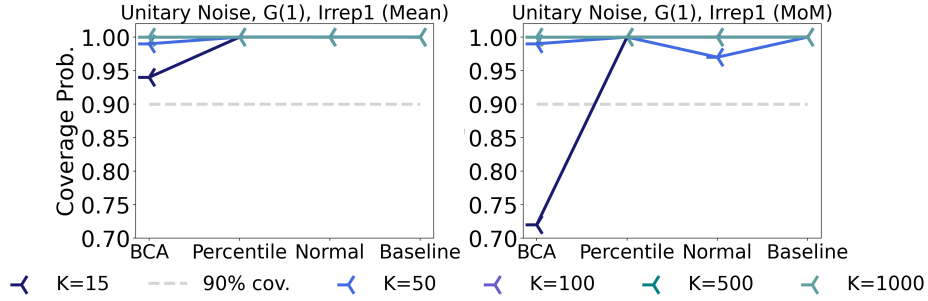


FIG. 13. Coverage probability for different confidence intervals and different sample sizes. The results are obtained by simulating the gate-set shadow protocol 100 times, and recomputing the confidence intervals, using each of the listed methods: the BCa method, the percentile method, taking the normal distribution assumption, and a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm. In this case, the protocol uses the CNOT-dehedral group, $G = G(1)$, as gate-set and gate errors are simulated with a unitary channel, namely the two-qubit zz gate $R_{zz}(\theta)$, with $\theta = 0.1$. Recall that this group has two decay factors λ_i . The present case shows the coverage probability for λ_1 .

1. Gate error model for Section III A

To obtain the numerical results in Fig.(3), we modeled gate errors as follows:

1. Gate operations obtained from only applying single-qubit gates to the two-qubit system are perturbed by a Pauli noise channel [50] (see Fig.(15), panel 1):

$$\Lambda_{\text{Pauli}}(\rho) = \sum_{j=0}^{15} p_j P_j \rho P_j, \text{ with } P_j \in \{\mathbb{1}, X, Y, Z\}^{\otimes 2}. \quad (\text{D1})$$

The set of amplitudes p_j form a probability distribution over the Pauli operators.

2. Gate operations that are produced by a composition of single-qubit gates with a single CNOT gate are perturbed by a composition of unitary errors, modeled by the zz -gate $R_{zz}(\theta)$ (see Fig.(15), panel 2), with Pauli noise:

$$\Lambda_1 = R_{zz}(\theta) \circ \Lambda_{\text{Pauli}}, \text{ with } R_{zz}(\theta) = \begin{pmatrix} e^{-i\theta/2} & 0 & 0 & 0 \\ 0 & e^{i\theta/2} & 0 & 0 \\ 0 & 0 & e^{i\theta/2} & 0 \\ 0 & 0 & 0 & e^{-i\theta/2} \end{pmatrix}. \quad (\text{D2})$$

3. Gates that require two CNOT gates to implement are perturbed by applying the previous error rate twice and, likewise, gates that require three CNOT gates are perturbed by three consecutive applications of Λ_1 .

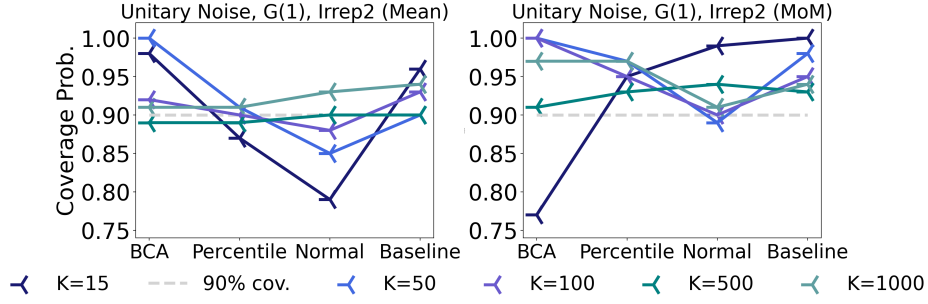


FIG. 14. Coverage probability for different confidence intervals and different sample sizes. The results are obtained by simulating the gate-set shadow protocol 100 times, and recomputing the confidence intervals, using each of the listed methods: the BCa method, the percentile method, taking the normal distribution assumption, and a two-standard deviation interval, where the standard deviation is retrieved from the covariance matrix of the fitting algorithm. In this case, the protocol uses the CNOT-dehdredal group, $G = G(1)$, as gate-set and gate errors are simulated with a unitary channel, namely the two-qubit zz gate $R_{zz}(\theta)$, with $\theta = 0.1$. Recall that this group has two decay factors λ_i . The present case shows the coverage probability for λ_2 .

We note that the $G(1)$ CNOT-dihedral group allows all of its elements to be built from gates requiring at most three CNOT gates. For the two-qubit Clifford group, we apply the same type of errors described above, following the decomposition of the group used in Ref. [51].

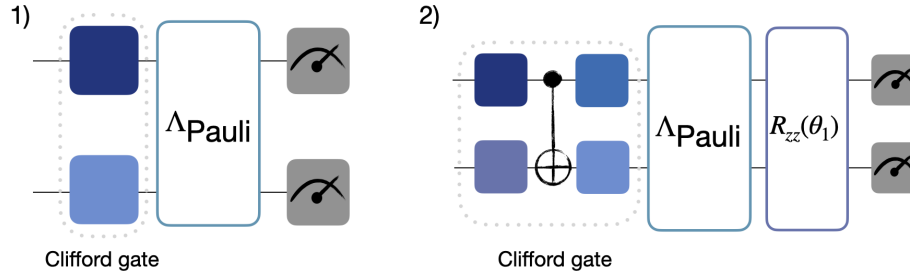


FIG. 15. Pictorial circuit-based representation of the gate errors used in simulating results in Section III A.

2. Gate error model for Section IV A

To obtain the numerical results in Fig.(5), we modeled operational crosstalk by letting the gate operations on qubit-1 induce an amplitude damping channel on qubit-2. Additionally, we modeled the noisy gate operations on qubit-1 by taking each operation to be distorted by a single qubit rotation $R_z(\theta_1)$. The angle θ_1 was fixed to $\theta_1 = 0.2$ for all gate operations. The error channel on the two-qubit system is then the result of applying two independent error channels to each qubit: a single-qubit z -rotation ($R_z(\theta_1)$) to qubit-1 and an amplitude damping channel on qubit-2 ($\Lambda_{AD}(\gamma_2)$). Each of these elementary channels are defined in the Kraus decomposition [7] as:

$$R_z(\theta_1)(\rho) = U_z(\theta_1) \rho U_z^\dagger(\theta_1), \quad U_z(\theta_1) = \begin{pmatrix} e^{-i\theta_1/2} & 0 \\ 0 & e^{i\theta_1/2} \end{pmatrix}, \quad (\text{D3})$$

$$\Lambda_{AD}(\gamma_2)(\rho) = \sum_{j=0}^1 K_j \rho K_j^\dagger, \quad K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma_2} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma_2} \\ 0 & 0 \end{pmatrix}.$$

A pictorial circuit-based representation of the resulting noisy sequence gates is shown in Fig.(16). The amplitude damping parameter γ_2 was fixed at $\gamma_2 = 0.02$. These settings produce the decaying curve in Fig.(5) for the gate-set $\mathcal{C}_1 \times \mathcal{I}$. The lower panel in the figure corresponds to evaluating the performance of the protocol across a range values for θ_1 , while keeping γ_2 unchanged.

The curve in Fig.(5) corresponding to the gate-set $\mathcal{C}_1^{\times 2}$ was generated by modeling the error gates in a similar fashion.

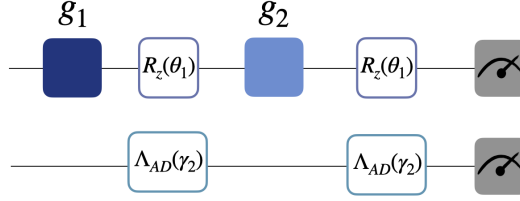


FIG. 16. Pictorial circuit-based representation of a set of two noisy gate operations from $\mathcal{C}_1 \times \mathcal{I}$ applied to a two-qubit system. The gate errors are modeled according to Eq.(D3). The ideal gate operations are represented by the gates g_1 and g_2 .

Since operations are now simultaneously applied to both qubits, we consider that operations on qubit-2 also induce a amplitude damping channel on qubit-1 ($\Lambda_{AD}(\gamma_1)$), while at the same time causing distorted gate operations on qubit-2 by the single-qubit rotation $R_z(\theta_2)$. Here, we set $\theta_1 = \theta_2 = 0.2$ and $\gamma_1 = \gamma_2 = 0.02$.

For both curves in the upper panel of Fig.(5), each sequence length results from an average of 1000 random gate sequences ($K=1000$).

Appendix E: Unital marginals reconstruction

Unital marginal channel reconstruction relies on the decomposition of the channel as a linear combination of elements of the local Clifford group, according to the formula [21]:

$$\Lambda_i = \frac{|\mathbb{P}_i|^2}{|\mathbb{C}_i|} \sum_{C \in \mathbb{C}_i} \lambda(C^\dagger, \Lambda) \mathbb{P}_i C \mathbb{P}_i, \quad (\text{E1})$$

$$\text{with } \lambda(C^\dagger, \Lambda) = \frac{\text{Tr}(\mathbb{P}_i C^\dagger \mathbb{P}_i \Lambda)}{|\mathbb{P}_i|}.$$

To ensure clarity of the presentation, let us expand on the adopted notation in Eq.(E1). We use $\mathbb{C}_i$ to denote the set of local Clifford gates acting non-trivially on the i th subspace. For example, for a two-qubit system, the set of operations $\{C_j \otimes \mathbb{1}\}_j$, with C_j a single-qubit Clifford operation, maps the full subspace spanned by the basis vectors $\{|\sigma_i \otimes \sigma_0\rangle\rangle\}$ onto itself. Applying this logic, and using the same labeling as in Tab. IV, we can identify $\mathbb{C}_1 = \{C_j \otimes \mathbb{1}\}_j$, $\mathbb{C}_2 = \{\mathbb{1} \otimes C_j\}_j$ and $\mathbb{C}_3 = \{C_j \otimes C_k\}_{(j,k) \neq 0}$. The dimension of each subspace is equivalent to $|\mathbb{P}_i| = \text{Tr}(\mathbb{P}_i)$. For any gate C in the set $\mathbb{C}_i$, its representative Pauli transfer matrix admits a block diagonal form. Hence, $\mathbb{P}_i C \mathbb{P}_i$ simply corresponds to extracting the block that acts on the i th subspace. Finally, $|\mathbb{C}_i|$ denotes the total number of elements in the set $\mathbb{C}_i$. The decomposition in Eq.(E1) is particularly convenient, because the expansion coefficients $\lambda(C^\dagger, \Lambda)$ are precisely the figures of merit estimated in the gate-set shadow protocol. Indeed, each $\lambda(C^\dagger, \Lambda)$ corresponds to taking the mean (or median-of-means) of the sequence correlation function f_A , with $A = C^\dagger$.

In Ref. [21], there are two proposed crosstalk metrics:

$$\delta\mathcal{R}_1 = |\Lambda_3 - \Lambda_1 \otimes \Lambda_2|, \quad \delta\mathcal{R}_2 = |(\Lambda_1 \otimes \Lambda_2)^{-1} \Lambda_3|. \quad (\text{E2})$$

In the main text, we have focused on $\delta\mathcal{R}_2$, but note that, for the most part, $\delta\mathcal{R}_1$ and $\delta\mathcal{R}_2$ are related. We note, however, that $\delta\mathcal{R}_2$ will be ill-defined if $\Lambda_1 \otimes \Lambda_2$ is not an invertible matrix. In Figs.(17)-(19), we illustrate the application of the two metrics to gauge correlations in the system, for different simulated error channels. The pathological case of the CNOT error [51] highlights the fact that the noise marginals do retain additional relevant information regarding the structure of the underlying gate noise. This can be understood as follows. Starting from the original gate noise, the construction of the unital marginals traces out all but information on the couplings between different elements within the same subspace. If now only information on the trace of the unital marginals is retained, then we operate at the level of the depolarizing-like noise, which is the effective noise accessed by the simultaneous RB protocol.

For all error models besides the CNOT case, the POVM elements were fixed to be the measurements in the z -computational basis, i.e. $\{E_x\}_x = \{|00\rangle\langle 00|, |01\rangle\langle 01|, |10\rangle\langle 10|, |11\rangle\langle 11|\}$. However, if we use the same set of measurements for the case where the average gate error is described by a CNOT channel, the resulting dataset cannot accurately retrieve structured information on the gate noise for all invariant subspaces. We may use the freedom in choosing the set of POVM operators and initial state to bypass this issue. Indeed, the numerical results in Fig.(19) employ a different choice for initial state and POVM elements, namely they make use of the initial state and POVM

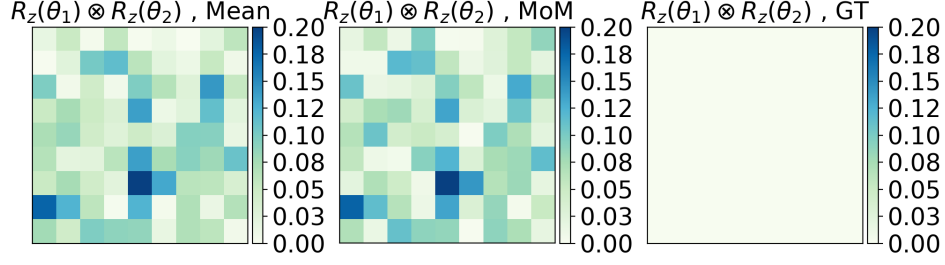


FIG. 17. $\delta\mathcal{R}_1$ for the case of a gate error model given by the tensor product of two single-qubit rotations along the z -axis. The rotational angles were fixed to be $\theta_1 = \theta_2 = 0.1$. From left to right: the first two figures are obtained by simulating the gate-set shadow protocol and reconstructing the unital marginals via Eq.(E1). The third figure corresponds to the actual value of $\delta\mathcal{R}_1$ (ground truth (GT)), derived using full knowledge on the simulated error model. Consistent with the fact that the simulated gate noise is uncorrelated, the matrix elements are small in amplitude, indicating that the resulting noise is close to uncorrelated noise. The data was simulated using 15000 random sequences per sequence length.

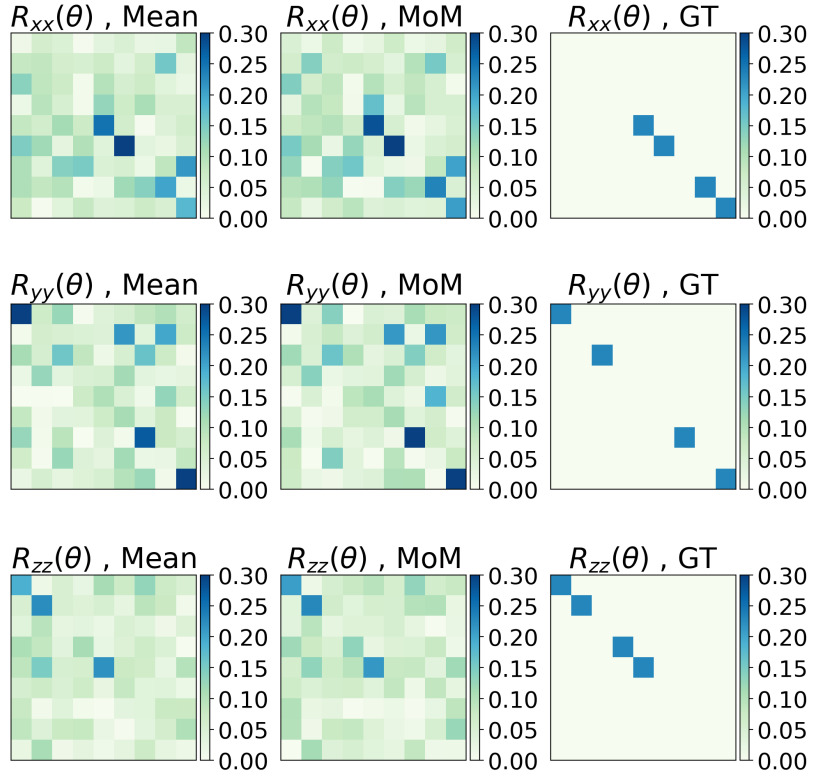


FIG. 18. $\delta\mathcal{R}_1$ for the case where the gate error model corresponds to one of the following three two-qubit rotation gates: $R_{xx}(\theta)$ (top panel), $R_{yy}(\theta)$ (middle panel) and $R_{zz}(\theta)$ (bottom panel). For all three cases, the angle θ was fixed at $\theta = 0.5$. In each row, from left to right: the first two figures are obtained by simulating the gate-set shadow protocol and reconstructing the unital marginals via Eq.(E1); the third figure corresponds to the actual value of $\delta\mathcal{R}_1$ (ground truth (GT)), derived using full knowledge on the simulated error model. The data was simulated using 15000 random sequences per sequence length.

elements described in Eq.(B25). The CNOT case, although arguably an unrealistic example, highlights the interplay between the performance of the gate-set shadow protocol and the choice of input and measurement basis.

In the main text, we also presented the conversion of the unital marginal channels into Pauli channels. Obtaining the Pauli channel can be done by computing the average of Λ_F over the set of all 2-qubit Pauli operations, in postprocessing. In the PTM formalism, taking the average of a given quantum operation $\mathcal{E}$, over all n -qubit Pauli gates, is

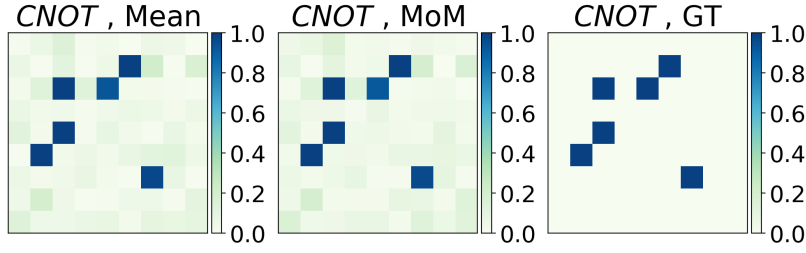


FIG. 19. $\delta\mathcal{R}_1$ for the case where the gate error model corresponds to a CNOT operation. From left to right: the first two figures are obtained by simulating the gate-set shadow protocol and reconstructing the unital marginals via Eq.(E1); the third figure corresponds to the actual value of $\delta\mathcal{R}_1$ (ground truth (GT)), derived using full knowledge on the simulated gate noise. The data was simulated using 15000 random sequences per sequence length.

given by:

$$\mathcal{E}_P = \frac{1}{4^n} \sum_{i=0}^{4^n-1} P_i \mathcal{E} P_i, \quad (\text{E3})$$

The motivation behind this transformation is as follows: since each of the unital marginals already corresponds to a projection of the gate noise to a particular subspace, we expect the effective Pauli noise produced by Eq.(E3) to reflect the dominant weight and locality features in its Pauli error rates. To retrieve the Pauli error rates, some additional steps are required. Firstly, we note that these are a probability distribution over the 2-qubit Pauli matrices. Since we are using the PTM formalism, we need to define a mapping between these two equivalent noise descriptions, i.e. from 2-qubit Pauli matrices to their vectorized versions. This is achieved through the relation [19]:

$$p_a = \frac{1}{4^n} \sum_{b=0}^{4^n-1} (-1)^{\langle P_a, P_b \rangle} (\mathcal{E}_P)_{b,b}, \quad (\text{E4})$$

where $\langle P_a, P_b \rangle = 0$ if the Pauli matrices P_a and P_b commute, and $\langle P_a, P_b \rangle = 1$ if they anti-commute. Imprecision in the reconstruction of the unital channel marginals can cause some of the estimated error rates p_a to be negative, particularly if their *true* value is zero. To reinforce the constraint that the amplitudes p_a should correspond to a probability distribution, we perform a final Euclidean projection onto the probability simplex [52]. After this final step, the resulting set of amplitudes $\{p_a\}_a$ is guaranteed to satisfy the constraint $\sum_a p_a = 1$, $p_a \geq 0$, although inaccuracies in the channel marginal reconstruction may still lead to a final probability distribution that differs from the ground truth one. A similar imprecision is present for the correlated RB coefficients. In particular, inaccuracies in the estimates for the λ parameters can cause small negative amplitudes for some of the estimated $\{\hat{\varepsilon}\}$ set, particularly when the corresponding ground truth amplitude is zero. In simulations, we first evaluate $\hat{\varepsilon}_2$. Since the amplitudes $\{\hat{\varepsilon}\}$ should be strictly positive [19], we have set a threshold value, where if $\hat{\varepsilon}_2$ is negative but has a small amplitude, i.e. $|\hat{\varepsilon}_2| < 0.1$, we fix it to be zero, and proceed similarly for the remaining $\hat{\varepsilon}$ parameters. This threshold value explains why not all amplitudes in Tab. III have a lower error estimate. In Tab. III, the error estimates were computed by considering how the set $\{\hat{\varepsilon}\}$ changed when the simultaneous RB decay rates increased or decreased, based on their respective error estimates. The simultaneous RB error estimates were defined from the fit uncertainty: they were retrieved from the standard deviation of the estimated covariance matrix (namely, we used 2σ standard errors).

-
- [1] S.J. Devitt, W.J. Munro and K. Nemoto, *Quantum error correction for beginners*, *Rep. Prog. Phys.* **76** (2013) 076001.
 - [2] C. Ryan-Anderson, J.G. Bohnet, K. Lee, D. Gresh, A. Hankin, J.P. Gaebler et al., *Realization of real-time fault-tolerant quantum error correction*, *Phys. Rev. X* **11** (2021) 041058.
 - [3] R. Acharya, D.A. Abanin, L. Aghababaie-Beni, I. Aleiner, T.I. Andersen, M. Ansmann et al., *Quantum error correction below the surface code threshold*, *Nature* **638** (2025) 920.
 - [4] J. Preskill, *Fault-tolerant quantum computation*, *quant-ph/9605011*.
 - [5] E. Knill, R. Laflamme and W.H. Zurek, *Resilient quantum computation: error models and thresholds*, *Proc. R. Soc. Lond. A* **454** (1998) 365.
 - [6] X. Yang, J. Chu, Z. Guo, W. Huang, Y. Liang, J. Liu et al., *Coupler-assisted leakage reduction for scalable*

- quantum error correction with superconducting qubits, *Phys. Rev. Lett.* **133** (2024) 170601.
- [7] I.L. Chuang and M.A. Nielsen, *Prescription for experimental determination of the dynamics of a quantum black box*, *J. Mod. Opt.* **44** (1997) 2455.
- [8] G.M. D'Ariano and P. Lo Presti, *Quantum Tomography for Measuring Experimentally the Matrix Elements of an Arbitrary Quantum Operation*, *Phys. Rev. Lett.* **86** (2001) 4195.
- [9] G. Torlai, C. Wood, A. Acharya, G. Carleo, J. Carrasquilla and L. Aolita, *Quantum process tomography with unsupervised learning and tensor networks*, *Nat. Commun.* **14** (2023) 2858.
- [10] C.H. Baldwin, A. Kalev and I.H. Deutsch, *Quantum process tomography of unitary and near-unitary maps*, *Phys. Rev. A* **90** (2014) 012110.
- [11] E. Magesan, J.M. Gambetta and J. Emerson, *Scalable and robust randomized benchmarking of quantum processes*, *Phys. Rev. Lett.* **106** (2011) 180504.
- [12] E. Magesan, J.M. Gambetta and J. Emerson, *Characterizing quantum gates via randomized benchmarking*, *Phys. Rev. A* **85** (2012) 042311.
- [13] J. Helsen, I. Roth, E. Onorati, A. Werner and J. Eisert, *General Framework for Randomized Benchmarking*, *PRX Quantum* **3** (2022) 020357.
- [14] E. Magesan, J.M. Gambetta, B.R. Johnson, C.A. Ryan, J.M. Chow, S.T. Merkel et al., *Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking*, *Phys. Rev. Lett.* **109** (2012) 080505.
- [15] R. Harper and S.T. Flammia, *Estimating the fidelity of T gates using standard interleaved randomized benchmarking*, *Quantum Sci. Technol.* **2** (2017) 015008.
- [16] J. Wallman, C. Granade, R. Harper and S.T. Flammia, *Estimating the coherence of noise*, *New J. Phys.* **17** (2015) 113020.
- [17] B. Dirkse, J. Helsen and S. Wehner, *Efficient unitarity randomized benchmarking of few-qubit Clifford gates*, *Phys. Rev. A* **99** (2019) 012315.
- [18] J.M. Gambetta, A.D. Córcoles, S.T. Merkel, B.R. Johnson, J.A. Smolin, J. Chow et al., *Characterization of Addressability by Simultaneous Randomized Benchmarking*, *Phys. Rev. Lett.* **109** (2012) 240504.
- [19] D.C. McKay, A.W. Cross, C.J. Wood and J.M. Gambetta, *Correlated Randomized Benchmarking*, 2003.02354.
- [20] H.Y. Huang, R. Kueng and J. Preskill, *Predicting many properties of a quantum system from very few measurements*, *Nat. Phys.* **16** (2020) 1050.
- [21] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A.H. Werner, J. Eisert et al., *Shadow estimation of gate-set properties from random sequences*, *Nat. Commun.* **14** (2023) 5039.
- [22] A. Silva and E. Greplova, *Hands-on introduction to randomized benchmarking*, *SciPost Phys. Lect. Notes* **97** (2025) .
- [23] A.W. Cross, E. Magesan, L. Bishop, J.A. Smolin and J.M. Gambetta, *Scalable randomised benchmarking of non-Clifford gates*, *npj Quantum Inf.* **2** (2016) 16012.
- [24] J. Hines, M. Lu, R.K. Naik, A. Hashim, J.-L. Ville, B. Mitchell et al., *Demonstrating scalable randomized benchmarking of universal gate sets*, *Phys. Rev. X* **13** (2023) 041030.
- [25] X. Xue, T.F. Watson, J. Helsen, D.R. Ward, D.E. Savage, M.G. Lagally et al., *Benchmarking gate fidelities in a Si/SiGe two-qubit device*, *Phys. Rev. X* **9** (2019) 021011.
- [26] J. Helsen, X. Xue, L. Vandersypen and S. Wehner, *A new class of efficient randomized benchmarking protocol*, *npj Quantum Inf.* **5** (2019) 71.
- [27] J. Claes, E. Rieffel and Z. Wang, *Character Randomized Benchmarking for Non-Multiplicity-Free Groups With Applications to Subspace, Leakage, and Matchgate Randomized Benchmarking*, *PRX Quantum* **2** (2021) 010351.
- [28] A. Silva and E. Greplova, "Every Benchmark All at Once." Accompanying code can be found here: <https://gitlab.com/QMAI/papers/everybenchmarkallatonce>.
- [29] J.M. Chow, J.M. Gambetta, A.D. Córcoles, S.T. Merkel, J.A. Smolin, C. Rigetti et al., *Universal Quantum Gate Set Approaching Fault-Tolerant Thresholds with Superconducting Qubits*, *Phys. Rev. Lett.* **109** (2012) 060501.
- [30] D. Greenbaum, *Introduction to Quantum Gate Set Tomography*, 1509.02921.
- [31] G. Lugosi and S. Mendelson, *Mean Estimation and Regression Under Heavy-Tailed Distributions: A Survey*, *Found. Comput. Math* **19** (2019) 1145.
- [32] S. Kimmel, M.P. da Silva, C.A. Ryan, B.R. Johnson and T. Ohki, *Robust Extraction of Tomographic Information via Randomized Benchmarking*, *Phys. Rev. X* **4** (2014) 011050.
- [33] A. Carignan-Dugas, J.J. Wallman and J. Emerson, *Bounding the average gate fidelity of composite channels using the unitarity*, *New J. Phys.* **21** (2019) 053016.
- [34] S. Garion and A.W. Cross, *Synthesis of CNOT-Dihedral circuits with optimal number of two qubit gates*, *Quantum* **4** (2020) 369.
- [35] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, vol. 57, Monographs on statistics and applied probability (1993).
- [36] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge Univ. Press) (1997).
- [37] S.G.J. Philips, M.T. Madzik, S.V. Amitonov, S.L. Snoo, M. Russ, N. Kalhor et al., *Universal control of a six-qubit quantum processor in silicon*, *Nature* **609** (2022) 919.
- [38] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen and R. Blume-Kohout, *Detecting crosstalk errors in quantum information processors*, *Quantum* **4** (2020) 321.
- [39] R. Klesse and S. Frank, *Quantum Error Correction in Spatially Correlated Quantum Noise*, *Phys. Rev. Lett.* **95** (2005) 230503.
- [40] A.G. Fowler, *Coping with qubit leakage in topological codes*, *Phys. Rev. A* **88** (2013) 042308.
- [41] C. Zhang, X.-C. Yang and X. Wang, *Leakage and sweet spots in triple-quantum-dot spin qubits: A molecular-orbital study*, *Phys. Rev. A* **97** (2018) 042326.
- [42] C.J. Wood and J.M. Gambetta, *Quantification and characterization of leakage errors*, *Phys. Rev. A* **97** (2018) 032306.
- [43] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, England (2010).

- [44] M. Tinkham, *Group theory and quantum mechanics*, Dover, New York (1964).
- [45] W. Fulton and J. Harris, *Representation theory: a first course*, vol. 129, Springer Science and Business Media (2013).
- [46] D. Gross, K. Audenaert and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs*, *J. Math. Phys.* **48** (2007) 052104.
- [47] A.A. Mele, *Introduction to Haar Measure Tools in Quantum Information: A Beginner's Tutorial*, *Quantum* **8** (2024) 1340.
- [48] C.J. Wood, J.D. Biamonte and D.G. Cory, *Tensor networks and graphical calculus for open quantum systems*, *Quant. Inf. Comp.* **15** (2015) 759–811.
- [49] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith et al., *PennyLane: Automatic differentiation of hybrid quantum-classical computations*, 1811.04968.
- [50] R. Harper, W. Yu and S.T. Flammia, *Fast Estimation of Sparse Quantum Noise*, *PRX Quantum* **2** (2021) 010322.
- [51] A.D. Córcoles, J.M. Gambetta, J.M. Chow, J.A. Smolin, M. Ware, J. Strand et al., *Process verification of two-qubit quantum gates by randomized benchmarking*, *Phys. Rev. A* **87** (2013) 030301.
- [52] W. Wang and M. Carreira-Perpinán, *Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application*, 1309.1541.