Randomized Benchmarking for Individual Quantum Gates

Onorati, E.; Werner, A. H.; Eisert, J.

Published in:
Physical Review Letters

DOI:
10.1103/PhysRevLett.123.060501

Publication date:
2019

Document version
Peer reviewed version

,

Randomized benchmarking for individual quantum gates

E. Onorati,[1] A. H. Werner,[2] and J. Eisert[1,3]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*Department of Mathematical Sciences, University of Copenhagen, DK-2100 København, Denmark*
[3]*Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany*

Any technology requires precise benchmarking of its components, and the quantum technologies are no exception. Randomized benchmarking allows for the relatively resource economical estimation of the average gate fidelity of quantum gates from the Clifford group, assuming identical noise levels for all gates, making use of suitable sequences of randomly chosen gates. In this work, we report significant progress on randomized benchmarking, by showing that it can be applied individually on a broad class of quantum gates outside the Clifford group , even for varying noise levels per quantum gate. This is possible at little overhead of quantum resources, but at the expense of a significant classical computational cost. At the heart of our analysis is a representation-theoretic framework which we bring into contact with classical estimation techniques based on bootstrapping and matrix pencils. We demonstrate the functioning of the scheme at hand of benchmarking tensor powers of T-gates. Apart from its practical relevance, we expect this insight to be relevant as it highlights the role of assumptions made on unknown noise processes when characterizing quantum gates at high precision.

Recent years have seen a rapid development in the precise preparation and manipulation of quantum states. Much of these advances have been triggered by the progress in the quantum technologies, anticipating computational quantum devices that outperform classical computers [1]. The increased sophistication of these experiments, however, is accompanied by an increased complexity of the task of verifying that the process has actually been implemented as expected. Unwanted and usually largely uncharacterized quantum noise hampers the correct functioning of quantum devices. Without the knowledge of such noise processes, yet, it is unclear how to achieve actionable advice on how to improve the experiment or whether the experiment has worked at all. The task of certifying quantum processes has hence been identified, unsurprisingly, as a key bottleneck in the field.

There are several strategies that can in principle be employed to tackle the problem. They can largely be classified in the assumptions they make about the given unknown process and the structure they employ. *Quantum process tomography* [2, 3] is an option, in principle, in that based on measurement data alone, an unknown process can be reconstructed. However, without any further structure or assumptions the scaling of the effort of quantum process tomography is daunting: for a generic d-dimensional quantum system, $\Theta(d^4)$ expectation values must be estimated to great accuracy. Methods exploiting structure, such as process tomography based on compressed sensing [4, 5] or diamond norm minimization [6] reduce the effort significantly. Also, the dynamics of quantum systems for short times can often be kept track of using tensor network states [7, 8]. Nevertheless, the benchmarking of large quantum circuits involving a large number of gates is out of scope in either of these approaches. This challenge is aggravated by the fact that gate errors suitable for *fault tolerant quantum computing* [1, 9–11] are extremely small, so that the characterization needs to be very precise.

Randomized benchmarking [12–18] – introduced in Ref. [12] – takes a different, more pragmatic route altogether. It acknowledges that full tomographic knowledge may be too much to ask for and achieves estimates of quantum gate errors making use of long sequences of randomly chosen Clifford gates. This is done to estimate a single quantity, the *average gate fidelity*, for a set of operations being the *Clifford group* in a large portion of the literature. That is to say, by making stronger assumptions about the underlying operations, one can go a long way reliably characterizing noise levels of quantum gates. The basic idea has been generalised in several ways; in particular, it has been shown that one does not need to employ an exact so-called *2-design* for randomized benchmarking [13, 19, 20], using instead, for instance, the single-qubit dihedral group [21]. To allow for extraction of the fidelity of a specific gate, a scheme called *interleaved randomized benchmarking* [22, 23] makes use of random sequences of Clifford gates interlaced by the particular gate whose noise is to be individually characterized, but it is still limited to the Clifford group itself or the T-gate. For protocols based on twirls over 2-design gates sets, such as the Clifford group, it has been recently shown [24] that they allow for strongly gate-dependent noise levels. However, other schemes which are able to go beyond the 2-design requirement assume the noise to be gate-independent [25], or weakly dependent [26]. In several ways the commonly made assumptions on the underlying processes are rather strong. For protocols not relying on 2-designs property, the premise on uniform noise channel for each gate of the group may be overly demanding; even if we take into consideration small generating gates sets, it is difficult to postulate that the noise of a single gate is equal to the one of a product of several consecutive operators. Here, we will see that some of these assumptions can be relaxed at little additional cost, as far as quantum resources are concerned.

In this work, we introduce the notion of randomized benchmarking for individual quantum gates, one that may be part of single layers of gate arrays. This work is expected to be significant in two ways: technically speaking, we introduce methods able to benchmark quantum gates individually, including ones that are outside the Clifford group. Since schemes for universal quantum computing necessarily make use of such

gates, this seems an important step forward. At the same time, we do not require twirling over the full Clifford group or a 2-design, but only over a relatively small local symmetry group coupled with transpositions gates. It is indeed a favourable feature of our approach to make use of a small group in contrast to large gate sets of previously known schemes.

The novel idea in the present work is to exploit the symmetry of the quantum gate itself in an appropriate fashion and hence reduce the amount of quantum resources and computational effort. To achieve this goal, we harness advanced tools from representation theory to arrive at schemes that require similar physical operations, departing from the paradigm of 2-designs, but which can make predictions beyond known prescriptions. These ideas are uplifted to functioning schemes by making use of *estimation techniques* such as matrix pencils [27]. It is key to the new approach that the demands for classical estimation are higher, which we accommodate by introducing more sophisticated tools to this task. Using such methods, we demonstrate that meaningful and experimentally accessible preparations are sufficient to render the recovery feasible at hand of randomized benchmarking data.

More conceptually speaking, and putting this contribution into a broader context, we show that one can interpolate between common assumptions made when characterizing quantum processes: In other words, there is "room in the middle" between full quantum process tomography, which is largely assumption-free but comes along with daunting resource requirements, and conventional randomized benchmarking, which requires significantly less effort and is also robust against *state preparation and measurement (SPAM)* errors, while making strong assumptions. Complementing this mindset in a very different way, Refs. [5, 28] provide method to extract tomographic information making use standard randomized benchmarking protocols, while maintaining very large data levels. We believe that this conceptual insight into the ontology of assumptions when characterizing quantum processes subject to quantum noise is equally important.

Exploiting local symmetries in randomized benchmarking. In the following, we are going to describe a protocol that provides the *average gate fidelity* of the noisy channel Λ which characterizes the imperfect implementation $\widetilde{U}$ of a target ideal unitary gate U,

$$\mathbb{E}(\mathcal{F}_{\widetilde{\mathcal{U}},\mathcal{U}}) = \mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) \coloneqq \int_{\mathrm{Haar}} \mathrm{Tr}\left[|\phi\rangle\langle\phi|\ \Lambda(|\phi\rangle\langle\phi|)\right] \mathrm{d}\phi. \tag{1}$$

It is key to our method to explicitly exploit the local and permutation symmetries of U, allowing for a drastic reduction of the fitting parameters, and also to inherit robustness with respect to SPAM errors. In this way, fitting models of well-known randomized benchmarking protocols can be uplifted to this setting involving fewer assumptions.

Throughout this work, we consider quantum gates acting on n-qubit systems and we are interested in benchmarking the accuracy of their implementation in a quantum circuit making use of the protocol that we are going to explain in a later section. The method is particularly suitable for gates consisting of tensor products of local gates, thus admitting additional symmetries with respect to the exchange of qubit subsystems and hence it is applicable in those situations of single layers of local unitary quantum gates. For concreteness, we shall put emphasis on single layers of circuits whose gates consist of tensor compositions of the T-gate with other gates belonging to the Clifford group $\mathcal{C}_n$, namely H, S and CNOT

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, \qquad T = \begin{pmatrix}\mathrm{e}^{-\mathrm{i}\pi/8} & 0\\ 0 & \mathrm{e}^{\mathrm{i}\pi/8}\end{pmatrix},$$
$$\mathrm{CNOT} = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}, \qquad S = \begin{pmatrix}1 & 0\\ 0 & \mathrm{i}\end{pmatrix}, \tag{2}$$

leading to a universal set. This setting is of paramount importance in state-of-the-art prescriptions of fault-tolerant quantum computing [1, 9–11]. More generally, the protocol is suitable to benchmark tensor products of local gates consisting in arbitrary rotations around one of the axes of the Bloch sphere. We will denote with $U = U_1 \otimes \cdots \otimes U_m$ the multi-qubit unitary operator – with U_j one or two qubits each – that we intend to benchmark, acting as a single layer of a unitary circuit. At the heart of the analysis will be its symmetry group, constructed from the symmetries of the local gates U_j composing U and the permutations of qubits which the local gate acts upon (cfr. Appendix D). More precisely, we choose the local symmetry group A_j of U_j as the subset of the single-qubit Clifford group, $\mathcal{C}_1$, whose elements commute with U_j. E.g., for $U_j = T = \exp(-\mathrm{i}\pi Z/8)$, the group of local symmetries is

$$A_T \coloneqq \{\, U \in \mathcal{C}_1 : U^\dagger Z U = Z \,\}. \tag{3}$$

This is an abelian group of 4 elements isomorphic to the cyclic group of order 4, $\mathbb{Z}_4$. The set of all possible permutations interchanging qubits affected by the same local gate is another symmetry group of the target unitary U; taking a pratical example, for the gate $U = T \otimes H \otimes T \otimes H \otimes T$, this group is isomorphic to $S_3 \times S_2$, i.e., all permutations of the first, third and fifth subsystems combined with the transposition of the second and forth subsystems. The full symmetry group G is then obtained through the *semi-direct product* $A_n \rtimes \Pi$, where A_n is the *direct product* of the local symmetry groups A_j constructed by the Kronecker product of the respective elements, and Π is the representation of the subgroup of S_n consisting of all allowed permutations of the qubits subsystems.

Role of abelian groups. In order to apply our full protocol and combine the group A_n with Π, all local symmetry groups A_j must be abelian. This is indeed a necessary condition to reconstruct the irreducible representations of the full group G with the sole knowledge of the composing groups, as we will discuss in the appendix [29]. Fortunately, this is the case for the centralizer in $\mathcal{C}_1$ of the gates in Eq. (2) and all other rotations around one of the Bloch axes. Should the local symmetry groups not be all abelian, the protocol is still valid setting $G = A_n$, i.e., without considering permutation symmetries.

Assumptions and physical motivation. We denote with calligraphic letters the channel acting by gate conjugation on density operators, i.e., $\mathcal{U}(\rho) \coloneqq U^\dagger \rho\, U$ and the noisy implementation of the idealized gate channel $\mathcal{U}$ as $\widetilde{\mathcal{U}} \coloneqq \Lambda_{\mathcal{U}} \circ \mathcal{U}$,

i.e., we account for a gate-dependent error channel $\Lambda_{\mathcal{U}}$ whose average fidelity we want to characterize with the proposed protocol. As randomized benchmarking can be interpreted as a trade-off between the level of characterization of the noise channel and the amount of physical and computational resources needed, we will make the following assumption: the Pauli-Liouville representation (cfr. Appendix E) of the twirled noise channel, $\Lambda_{\mathcal{U}}^{G} \coloneqq |G|^{-1} \sum_{j \in \mathbb{N}_{|G|}} \mathcal{G}_j^{\dagger} \Lambda_{\mathcal{U}} \mathcal{G}_j$, is *almost jointly diagonalizable* with the target unitary channel $\mathcal{U}$ (where again we consider it as a matrix in Pauli-Liouville representation), in the spirit of Ref. [30]. This means that when the matrix $\mathcal{U}$ is brought to diagonal form by some unitary transformation V, the off-diagonal element of $\Lambda_{\mathcal{U}}^{G}$ under the same transformation are small and in particular, $\Lambda_{\mathcal{U}}^{G}$ will approximately leave the same subspaces invariant. This is true in two cases. The first possibility is that both $\mathcal{U}$ and the twirled noise $\Lambda_{\mathcal{U}}^{G}$ are diagonalizable simultaneously in a certain basis, e.g., when the decomposition of the representation of the symmetry group into irreducible representations has no multiplicity: in this case, both $\mathcal{U}$ and the twirled noise $\Lambda_{\mathcal{U}}^{G}$ are "forced" to assume a diagonal form with respect to the irreducible subspaces. If this is not the case, then $\Lambda_{\mathcal{U}}^{G}$ assumes a sparse form with some off-diagonal entries, which have to be small with respect to the diagonal elements. This is fulfilled whenever the original noise channel $\Lambda_{\mathcal{U}}$ related to the implementation of the gate $\mathcal{U}$ was almost jointly diagonalizable to begin with or, put in another perspective [30], it is almost commuting. This assumption is valid when the gate U is generated by a Hamiltonian H applied for some run time t [31], i.e., $U = \mathrm{e}^{-\mathrm{i}Ht}$, which can be perturbed for a small fraction of the time, or be applied for too much or too little time (please refer to Section F in the appendix for more details). Furthermore, we ask the gates belonging to the symmetry group G to be implementable with high accuracy (see Appendix I). These gates either perform a permutation of the subspaces of the system or belong to the Clifford group and so can be for instance benchmarked with the well-known methods [15, 16, 24].

The protocol. We propose a slightly modified version of the previous protocols. We apply in succession channels defined by the gate U after the one induced by a gate uniformly drawn at random from the symmetry group G. In addition, we can target the different symmetry subspaces that are stabilized by $\mathcal{G}$ by choosing an appropriate initial state through the application of projectors decomposing a density operator into basis vectors of distinct irreducible subspaces (see Ref. [32] and Appendix L, Section 2). Note that, unlike previous protocols, the target gate U is not part of the twirling group G: This is one of the reason why one can benchmark arbitrarily small rotations over the Bloch axes with a relatively small number of gates. For a fixed sequence length ℓ the protocol consists of the steps:

(i) Prepare an initial state ρ with support in the target invariant subspace(s).

(ii) Draw a random sequence $\mathbf{k}_\ell = (k_1, \ldots, k_\ell) \in \mathbb{N}_{|G|}^{\ell}$.

(iii) Apply the following operation generated by the symmetry operations $\mathcal{G}_{k_i}$ to the initial state ρ

$$\mathcal{C}_{\mathbf{k}_\ell}(\rho) = \mathcal{G}_{\mathrm{inv}} \circ \mathcal{U} \circ \mathcal{G}_{k_\ell} \circ \cdots \circ \mathcal{U} \circ \mathcal{G}_{k_2} \circ \mathcal{U} \circ \mathcal{G}_{k_1}, \qquad (4)$$

where $\mathcal{G}_{\mathrm{inv}} \coloneqq \mathcal{G}_{k_1}^{\dagger} \circ \cdots \circ \mathcal{G}_{k_\ell}^{\dagger}$ is the channel given by the composition of the inversion of all previous random gates channels.

(iv) Perform a suitable POVM E, e.g., by applying the projector used to construct the input state ρ according to Eq. (21) in the appendix, and measure the *survival probability* $F_{\mathbf{k}_\ell} = \mathrm{Tr}[E\, \mathcal{C}_{\mathbf{k}_\ell}(\rho)]$. To obtain an appropriate precision for $F_{\mathbf{k}_\ell}$, this step has to be repeated sufficiently often.

(v) Repeat the previous step for sufficiently many (say N) random sequences $\mathbf{k}_{\ell,1}, \ldots, \mathbf{k}_{\ell,N}$ of length ℓ. Then, calculate the *sequences survival probability*

$$F_{\mathrm{seq}}(\ell, \rho) = \frac{1}{N} \sum_{\mathbf{k}_\ell} F_{\mathbf{k}_\ell} = \frac{1}{N} \sum_{\mathbf{k}_\ell} \mathrm{Tr}[E\, \mathcal{C}_{\mathbf{k}_\ell}(\rho)]. \qquad (5)$$

The number N of random sequences should be chosen such that

$$F_{\mathrm{seq}} \approx F_{\mathrm{avg}}\ , \qquad (6)$$

where F_{avg} is the survival probability averaged over all possible sequences. The choice can be motivated by an analysis on the variance of the random variable F, with $F_{\mathbf{k}_\ell}$ being a *realization* and F_{avg} the *mean* of the distribution (see Appendix J for a bound on the number of samples). Note that, for a Clifford circuit, one requires a "relatively small" number of sequences to approximate F_{avg} [17, 33].

(vi) Repeat the previous steps for different lengths ℓ.

(vii) Insert F_{seq} into the zeroth-order fitting model (whose derivation is discussed in the next paragraph),

$$F_{\mathrm{avg}}^{(0)}(\ell, \rho) = \sum_j (\lambda_j\, d_j)^{\ell} \xi_j \qquad (7)$$

where the sum runs over the eigenvalues $\{\, d_j \,\}_j$ of the target matrix $\mathcal{U}$ belonging to the space which ρ has support over, with $\xi_j \coloneqq \mathrm{Tr}[E\, \Lambda'(v_j)]\, \langle \rho, v_j \rangle$ absorbing the state preparation and measurement errors and where $\{\, v_j \,\}_j$ is the set of the basis vectors diagonalizing $\mathcal{U}$. The set $\{\, \lambda_j \,\}_j$ corresponds to the diagonal elements of Λ with respect to this basis.

(viii) Subsequently, we retrieve the parameters $\{\lambda_j\}$, which characterize the average gate fidelity of $\widetilde{\mathcal{U}}$ with respect to $\mathcal{U}$ (cfr. Appendix H) according to the relation

$$\mathbb{E}(\mathcal{F}_{\Lambda^G, \mathcal{I}}) = \frac{\sum \lambda_j + d}{d(d+1)}, \qquad (8)$$

using classical estimation techniques.

The fitting model. Considering a noise channel Λ (we now drop the subscript $\mathcal{U}$ to lighten notation) at each implementation of $\mathcal{U} \circ \mathcal{G}_k$, we can write

$$\mathcal{C}_{k_\ell} = \Lambda' \circ \mathcal{G}_{\mathrm{inv}} \bigcirc_{t=\ell}^{1} \Lambda \circ \mathcal{U} \circ \mathcal{G}_{k_t}. \qquad (9)$$

Note that the error channel Λ' characterizing the implementation of $\mathcal{G}_{\mathrm{inv}}$ can be different from the error for the implementation of $\mathcal{U} \circ \mathcal{G}_{k_t}$. Now, defining recursively $\mathcal{B}_{k_t} \coloneqq \mathcal{G}_{k_t} \circ \mathcal{B}_{k_{t-1}}$

with $\mathcal{B}_{k_1} = \mathcal{G}_{k_1}$, and using the invariance of U under the action of G, we can rewrite

$$\mathcal{C}_{k_\ell} = \Lambda' \bigcirc_{t=\ell}^{1} \mathcal{B}_{k_t}^{\dagger} \circ \Lambda \circ \mathcal{B}_{k_t} \circ \mathcal{U}. \tag{10}$$

When averaging over all possible sequences, we get $F_{\text{avg}}(\ell, \rho) = \text{Tr}[E\, \mathcal{C}_{\text{avg}}(\rho)]$ with $\mathcal{C}_{\text{avg}} = \Lambda' \bigcirc_{t=\ell}^{1} \Lambda^G \circ \mathcal{U}$, where $\Lambda^G \coloneqq |G|^{-1} \sum_{j \in \mathbb{N}_{|G|}} \mathcal{B}_j^{\dagger} \circ \Lambda \circ \mathcal{B}_j$ is now like U invariant with respect to the action of G. At this point we know that by Schur's Lemma (see Appendix C) the Pauli-Liouville representations of $\mathcal{U}$ and Λ^G are block-diagonalizable. If the decomposition of G does not contain any multiplicity for all distinct irreducible subspaces, the two matrices are simultaneously diagonalizable. If conversely multiple of the same irreducible representations occurs, in general there is no basis which brings both into a diagonal form and so, when diagonalizing $\mathcal{U}$, the matrix representation of Λ^G will assume a block form, where each of these blocks corresponds to an irreducible representation. For the zeroth-order model, we consider the diagonal elements of Λ^G only, and in particular the set $\{\, \lambda_j \,\}_j$ accounted in Eq. (7) denotes the ones belonging to the support of ρ. Let us write $\Lambda^G = \Lambda_0 + \Lambda_{\text{off}}$, with Λ_0 being jointly diagonalizable with $\mathcal{U}$ and Λ_{off} having smaller matrix entries due to the previously discussed assumption. Then, in zeroth-order, we have $F_{\text{avg}}^{(0)}(\ell, \rho) = \text{Tr}[E\, \Lambda' \left(\bigcirc_{t=\ell}^{1} \Lambda_0 \circ \mathcal{U}\right)(\rho)]$, from which Eq. (7) follows. If one needs to take into consideration off-diagonal entries of Λ^G too, a first-order model is necessary (see Appendix G).

Statistical analysis and bootstrapping. In order to estimate the average fidelity with the help of Eq. (8) we need to extract the decay rates $\{\, \lambda_j \,\}_j$ from the measured data. According to the protocol, this data consists of the survival probabilities $\{F_{\mathbf{k}_{\ell,q}}\}$ measured for $q = 1, \ldots, N$ randomly chosen sequences of lengths $\ell = 1, \ldots, \ell_{\max}$. Let us denote by $F_{\text{seq}}(\ell)$ the average of $F_{\mathbf{k}_{\ell,q}}$ with respect to the randomly chosen symmetry-group sequences $k_{\ell,q}$ for fixed sequence length ℓ. The zeroth-order fitting model from Eq. (7) then tells us that $F_{\text{seq}}(\ell) \approx \sum_j (\lambda_j d_j)^{\ell} \xi_j$ and accordingly, we can extract the parameters $\{\, \lambda_j \,\}_j$ from the sequence $(F_{\text{seq}}(\ell))_{\ell=1}^{\ell_{\max}}$ with the help of matrix pencil based signal reconstruction (see Appendix L). In order to improve our estimates we combine this approach with a bootstrapping procedure in the following way: To obtain a single bootstrap sample, we choose from the complete set of measured survival probabilities $\{F_{\mathbf{k}_{\ell,q}}\}_{q=1,\ell=1}^{N,\ell_{\max}}$, for each sequence length ℓ a random subset of the sequences $\mathbf{k}_{\ell,q}$ for which we then compute the average $F_{\text{seq}}(\ell)$ and extract $\{\, \lambda_j \,\}_j$ as described before. Repeating this process for different random samples of the $\mathbf{k}_{\ell,q}$, we obtain different estimates for the average fidelity, which is then averaged again with respect to the bootstrap samples in order to produce our final estimate for the average fidelity.

Recovery for single T-gate and two T-gates. In order to check the performance of our proposed protocol, we consider two examples: A single T-gate and the tensor product of two T-gates. In the single T-gate case, the generating Hamilonian is perturbed by an additional σ_x term and we write $H = \frac{\pi}{8}\sigma_z - \varepsilon\sigma_x$, whereas in the two T-gates we consider $H = \frac{\pi}{8}\left(\sigma_z \otimes \mathbb{1} + \mathbb{1} \otimes \sigma_z - \varepsilon\, \sigma_x \otimes \sigma_x\right)$ for varying perturbation strength ε. Fig. 1 compares the actual values of the average fidelity $\mathbb{E}(\mathcal{F})$, with the ones retrieved by fitting the zeroth-order model for different values of ε. The details of the fitting procedure are described in Appendix M. From the two graphs it can be seen that we obtain very accurate predictions for the average fidelity already in the regime $\varepsilon \approx 0.25$.

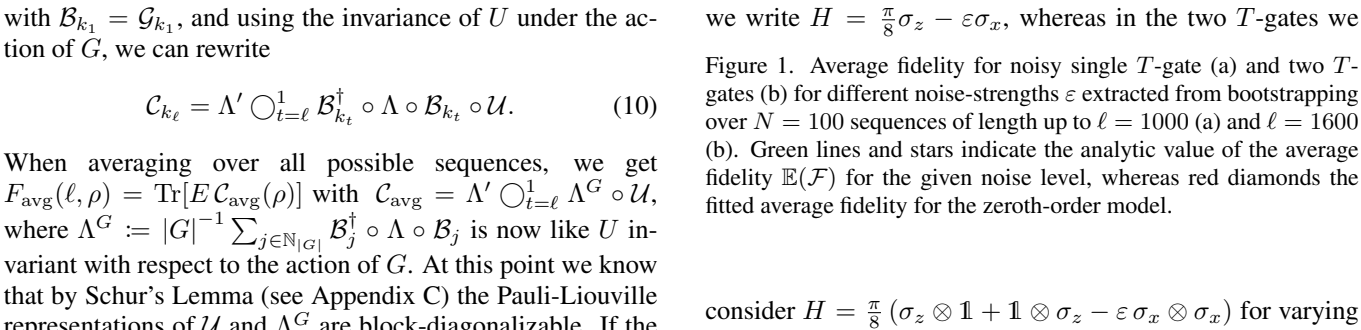

Figure 1. Average fidelity for noisy single T-gate (a) and two T-gates (b) for different noise-strengths ε extracted from bootstrapping over $N = 100$ sequences of length up to $\ell = 1000$ (a) and $\ell = 1600$ (b). Green lines and stars indicate the analytic value of the average fidelity $\mathbb{E}(\mathcal{F})$ for the given noise level, whereas red diamonds the fitted average fidelity for the zeroth-order model.

Conclusions and outlook. In this work, we have proposed a novel approach to randomized benchmarking, developing a protocol that allows to estimate individually the average gate fidelity of a unitary operator whose imperfect implementation is described by a gate-dependent noise channel, hence lifting some of the weaknesses related to the well-known protocols while maintaining robustness against SPAM errors. The key idea involves the twirl of the noise channel over the symmetry group of the target gate given by the composition of local symmetries with respect to the qubit subsystems and permutations thereof: This reduces considerably the amount of parameters that one has to retrieve and also the number of random gates needed. We have produced numerical simulations showing the validity of the protocol for a single-qubit and two-qubit T-gate, which proves to be faithful for relatively large noise levels. While we do not have an analytical formula for the scalability of the protocol, also due to the very wide range of possible gates that can be benchmarked in this way, the example of multiple tensor copies of the T-gate in Appendix K suggests that the number of non-linear parameters in the fitting model is drastically inferior compared to the full matrix dimension when increasing the number of qubits. Further investigations in this direction would be certainly of interest. Another open question concerns the number of different random samples needed to approximate the average fidelity with confidence, and in particular to obtain a bound in the fashion of Ref. [33]. With this letter, we hope to further stimulate research on randomized benchmarking outside the more established domain of the Clifford group and to explore the territory between randomized benchmarking and more traditional quantum process tomography.

Acknowledgements. We thank R. Kueng, O. Buerschaper, R. Di Candia, S. T. Flammia, and M. Wageringel for valuable discussions. This work has been supported by the ERC (TAQ), the DFG (EI 519/7-1, EI 519/11-1, CRC 183, DAEDALUS), the excellence cluster MATH+, and the Templeton Foundation. This work has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 817482 (PASQuanS). AHW thanks the Humboldt Foundation for its support with a Feodor Lynen Fellowship.

[1] A. Acin, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm. The European quantum technologies roadmap. *New J. Phys.*, 20:080201, 2018.

[2] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.*, 44:2455–2467, 1997.

[3] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, 2008.

[4] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, 2010.

[5] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch. Recovering quantum gates from few average gate fidelities. *Phys. Rev. Lett.*, 121:170502, 2018.

[6] M. Kliesch, R. Kueng, J. Eisert, and D. Gross. Improving compressed sensing with the diamond norm. *IEEE Trans. Inf. Th.*, 62:7445, 2016.

[7] A. Steffens, M. Friesdorf, T. Langen, B. Rauer, T. Schweigler, R. Hübener, J. Schmiedmayer, C. A. Riofrio, and J. Eisert. Towards experimental quantum field tomography with ultracold atoms. *Nature Comm.*, 6:7663, 2015.

[8] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos. Efficient tomography of a quantum many-body system. arXiv:1612.08000.

[9] J. Preskill. Fault-tolerant quantum computation. *arXiv:quant-ph/9712048*, 1997.

[10] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508, 2014.

[11] J. M. Chow, J. M. Gambetta, E. Magesan, S. J. Srinivasan, A. W. Cross, D. W. Abraham, N. A. Masluk, B. R. Johnson, C. A. Ryan, and M. Steffen. Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nature Comm.*, 5:4015, 2014.

[12] J. Emerson, R. Alicki, and K. Zyczkowski. Scalable noise estimation with random unitary operators. *J. Opt. B*, 7:S347–S352, 2005.

[13] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *arXiv:0606161*, 2006.

[14] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, 2008.

[15] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, 2012.

[16] E. Magesan, J. M. Gambetta, and J. Emerson. Robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:042311, 2011.

[17] J. J. Wallman and S. T. Flammia. Randomized benchmarking with confidence. *New. J. Phys.*, 16:103032, 2014.

[18] K. Boone, A. Carignan-Dugas, J. J. Wallman, and J. Emerson. Randomized benchmarking under different gatesets. *arXiv:1811.01920*.

[19] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.*, 48:052104, 2007.

[20] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta. Scalable randomised benchmarking of non-Clifford gates. *npj Quant. Inf.*, 2, 2016.

[21] C. A. Dugas, J. Wallman, and J. Emerson. Characterizing universal gate sets via dihedral benchmarking. *Phys. Rev. A*, 92:060302, 2015.

[22] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.*, 109:080505, 2012.

[23] R. Harper and S. T. Flammia. Estimating the fidelity of T gates using standard interleaved randomized benchmarking. *Quant. Sc. Tech.*, 2:015008, 2017.

[24] J. J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, 2018.

[25] D. S. Franca and A. K. Hashagen. Approximate randomized benchmarking for finite groups. *J. Phys. A*, 51:395302, 2018.

[26] W. G. Brown and B. Eastin. Randomized benchmarking with restricted gate sets. *Phys. Rev. A*, 97:062323, 2018.

[27] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38:814–824, May 1990.

[28] 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*, 4:011050, 2014.

[29] The supplements cite Refs. [15, 17, 27, 28, 33–43].

[30] K. Glashoff and M. M. Bronstein. Almost-commuting matrices are almost jointly diagonalizable, 2013. arXiv:1305.2135.

[31] B. E. Anderson, H. Sosa-Martinez, C. A. Riofrio, I. H. Deutsch, and P. S. Jessen. Accurate and robust unitary transformations of a high-dimensional quantum system. *Phys. Rev. Lett.*, 114:240401, 2015.

[32] J. Helsen, X. Xue, L. M. K. Vendersypen, and S. Wehner. A new class of efficient randomized benchmarking protocols, 2018. arXiv:1806.02048.

[33] J. Helsen, J. Wallman, S. T. Flammia, and S. Wehner. Multi-qubit randomized benchmarking using few samples, 2017. arXiv:1701.04299.

[34] J.-P. Serre. *Linear representations of finite groups*. Springer, 1996.

[35] R. Berndt. *Representations of linear groups*. Vieweg, first edition edition, 2007.

[36] F. Casas, A. Murua, and M. Nadinic. Efficient computation of the Zassenhaus formula. *Comp. Phys. Comm.*, 183:2386–2391, 2012.

[37] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Phys. Rev. A*, 303:249–252, 2002.

[38] D. P. O'Leary and B. W. Rust. Variable projection for nonlinear least squares problems. *Computational Optimization and*

Applications, 54:579–593, 2013.

[39] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Ac. Sp. Sig. Proc.*, 38:814–824, May 1990.

[40] T. K. Sarkar and O. Pereira. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *IEEE Ant. Prop. Mag.*, 37:48–55, Feb 1995.

[41] D. Potts and M. Tasche. Parameter estimation for nonincreasing exponential sums by prony-like methods. *Linear Algebra and its Applications*, 439:1024 – 1039, 2013. 17th Conference of the International Linear Algebra Society, Braunschweig, Germany, August 2011.

[42] J.-M. Papy, L. De Lathauwer, and S. Van Huffel. Common pole estimation in multi-channel exponential data modeling. *Signal Processing*, 86:846 – 858, 2006.

[43] D. S. Moore, G. P. McCabe, and B. A. Craig. *Introduction to the practice of statistics*. W. H. Freeman, 2016.

,

Supplemental material: Randomized benchmarking for individual quantum gates

E. Onorati,[1] A. H. Werner,[2] and J. Eisert[1,3]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*Department of Mathematical Sciences, University of Copenhagen, DK-2100 København, Denmark*
[3]*Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany*

Structure of the appendix

This appendix contains technical details for the scheme proposed in the letter and is structured as follows in order to have some logical flow and help readability. The first Sections A and B provide the basic theoretic elements for group and representation theory, whose most relevant tool to this work is separately given in Section C. As we mentioned in the main part of the letter, the core idea of our novel randomized benchmarking protocol is the coupling of local and permutation symmetries in order to produce a larger symmetry group of the target gate; the machinery for retrieving the irreducible characters of a semi-direct product group from the sole knowledge of the irreducible representations of its constituent subgroups is illustrated in Section D. After this first part, we focus on other particular aspects of the protocol. We justify in Section F the assumption on almost jointly diagonalizable matrix representations of the noise and ideal unitary channels, we discuss in Section G the derivation of a first-order fitting model to account for off-diagonal terms, in Section H we connect the retrieved parameters to the average gate fidelity, we describe how to separate this quantity for the target gate from the one of the twirling group in Section I and we provide in Section J a theoretical bound to the number of different random sequences needed to approximate the full sample average. The last part of the appendix is dedicated to a concrete example, where we apply our protocol to (up to four) tensor copies of the T-gate. We first make use of the machinery of Section D to construct the character tables of the symmetry groups in Section K. We then describe in Section L the recovery techniques for the parameters in the zeroth-order fitting model, combining the matrix pencil method with a bootstrapping approach. Finally, we provide the numerics for the fidelity estimation in the single and two T-gates cases in Section M.

A. Group theory

In this section, we provide the mathematical tools from group and representation theory which our benchmarking protocol relies upon; the reader already familiar with this mathematical field can skip the following introductory material.

Definition 1 (Group)**.** *A* group G *is a set of elements equipped with a binary operation satisfying the following properties:*
Closure*: For all* $g, h \in G,\ g \cdot h \in G$.
Associativity*: For all* $g, h, k \in G,\ (g \cdot h) \cdot k = g \cdot (h \cdot k)$.
Identity element*: There exist a unique identity element* e *such that, for all* $g \in G$, $e \cdot g = g \cdot e = g$.
Inverse*: for every element* $g \in G$ *there exist an inverse element* g^{-1} *such that* $g^{-1} \cdot g = g \cdot g^{-1} = e$.

If two groups can be linked by a group isomorphism, they are said to be *isomorphic*. They will then have many properties in common, in particular the same multiplication table and character table. One can therefore easily obtain information about a group if it is possible to find a group isomorphism connecting it to another well-known group; this is indeed what we do in our protocol to deal with the local symmetry groups. A subset $H \subset G$ is called *subgroup* if all above conditions are still satisfied, e.g., the subset is closed with respect to the group operation. A subgroup N such that $g^{-1} n\, g \in N$ for all $n \in N,\ g \in G$ is said to be *normal* and this is denoted by $N \lhd G$.

One can define a *(left) group action of* G *on a set* M by a function

$$\begin{aligned} \phi : \quad & G \times M \to M, \\ & (g, m) \mapsto \phi(g, m) \end{aligned} \tag{1}$$

that fulfills the following two axioms:
Identity: for all $m \in M,\ \phi(e, m) = m$,
Compatibility: for all $m \in M, g, h \in G\ \phi(g, \phi(h, m)) = \phi(g \cdot h, m)$. With this definition, we can furthermore define the following.

Definition 2 (Orbit)**.** *An* orbit $G.m$ *of an element* $m \in M$ *is given by all elements in* M *obtained by the action of* G*, i.e.,*

$$G.m \coloneqq \{\, \phi(g, m) : g \in G \,\}\,. \tag{2}$$

The action of G on M induces a *partition* of the set M itself, i.e., it regroups the elements into subsets such that every element $m \in M$ is contained in one and only one of these.

Definition 3 (Stabilizer subgroup)**.** *The* stabilizer subgroup *of* G *with respect to* m *is the set of all elements on* G *such that*

$$G_m \coloneqq \{\, g \in G : \phi(g, m) = m \,\}\,. \tag{3}$$

It is always possible to couple two groups to generate a new one. This is indeed what we are seeking, having to combine symmetry of the local gates together with the invariance with respect to (certain) permutations thereof.

Definition 4 (Direct product)**.** *Given two groups* G *and* H*, the* direct product $G \times H$ *is a* Cartesian product *of ordered pairs* (g, h)*, with* $g \in G, h \in G$ *equipped with a binary operation acting component-wise, that is*

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot g_2, h_1 \cdot h_2). \tag{4}$$

This new structure satisfies all axioms of closure, associativity, existence of identity – given by (e_G, e_H) – and inverse element– (g^{-1}, h^{-1}) being the inverse of (g, h) – and so it is a group. An alternative way to construct a new group from is given by the *semi-direct product*.

Definition 5 (Outer semi-direct product)**.** *Let N, H be groups, $\varphi : H \to \mathrm{Aut}(N)$ be a homomorphism from H to the set of automorphisms of N. Then the* (outer) semi-direct product with respect to ϕ*, denoted by $G = N \rtimes_\varphi H$, is the group whose underlying set are the pairs $(n, h) \in N \times H$ equipped with an operation defined as*

$$\bullet : G \times G \to G \tag{5}$$

$$\begin{aligned}((n_1, h_1), (n_2, h_2)) &\mapsto (n_1, h_1) \bullet (n_2, h_2) \\ &= (n_1 \cdot \varphi_{h_1}(n_2), h_1 \cdot h_2),\end{aligned} \tag{6}$$

where $n_1, n_2 \in N, h_1, h_2 \in H$.

This structure is again a group according to the defining axioms, with identity element (e_N, e_H) and inverse $(n, h)^{-1} = (\varphi_{h^{-1}}(n^{-1}), h^{-1})$. Note that the the set $\{(n, e_H) : n \in N\}$ is a normal subgroup of G isomorphic to N.

It is also possible to go the other way around and obtain from a group G and a normal subgroup N a new group called *quotient group*, denoted by G/N. This is the set of all cosets of N in G, i.e.,

$$G/N \coloneqq \{\, gN : g \in G \,\}, \tag{7}$$

where gN is the *left coset of N in G* , namely

$$gN \coloneqq \{\, gn : n \in N \,\}. \tag{8}$$

The latter definitions stands for all subgroup N, not necessarily normal, however when N is normal the left coset and the right coset (defined analogously) coincide. The set G/N is then a group under the operation $(gN) \cdot (hN) = (gh)N$. We conclude this paragraph with the following definition of the *canonical projection* which is involved in the construction of irreducible representations of a semi-direct product group, as we will see in the next subsection.

Definition 6 (Canonical projection)**.** *Let $N \lhd G$. The group homomorphism*

$$\tau : G \to G/N, \tag{9}$$

$$g \mapsto gN \tag{10}$$

is called canonical projection*.*

B. Representation theory

We are now going to introduce representations and some of the most important results involving them. The following propositions are the mathematical tools at the heart of our randomized benchmarking protocol.

Definition 7 (Representation)**.** *A* representation *of a group G on a vector space V is a group homomorphism onto the general linear group on V,i.e., a map*

$$\pi : H \to \mathrm{GL}(V), \tag{11}$$

$$g \mapsto \pi(g) \tag{12}$$

such that

$$\pi(g) \cdot \pi(h) = \pi(g \cdot h). \tag{13}$$

A representation is said to be *faithful* if it is injective, and its dimension corresponds to the dimension of the vector field V. A subspace $W \subset V$ is said to be *invariant* if, for all $g \in G$ and $w \in W$,

$$\pi(g)w \in W. \tag{14}$$

Furthermore, a representation is said to be *irreducible* if the only invariant subspaces are $\{\, 0 \,\}$ and V itself; often, this is abbreviated as *irrep*. Every complex representation of a finite group is *completely reducible*, i.e., it can be decomposed as a direct sum of irreducible representations. This property, together with Schur's Lemma, makes irreducible representations and their *characters* a central object in the theory and will also be particularly relevant in our work.

Definition 8 (Character of a representation)**.** *The* character χ_π of a representation π of a group G on V *is given by*

$$\chi_\pi(g) = \mathrm{Tr}[\pi(g)]. \tag{15}$$

The dimension of a representation corresponds then to its character at the identity element, $\chi_\pi(e)$. For finite group, the number of irreducible representations is again finite, and the following result is useful to check if all irreducible representations of a given group have been found.

Proposition 9 (Group order and irreducible representations dimension)**.** *The order of a group G and the dimension of its irreducible representations are linked by*

$$|G| = \sum_{\alpha \,:\, \pi_\alpha \mathrm{irrep}} \chi_{\pi_\alpha}(e)^2. \tag{16}$$

One of the most important properties for character of irreducible representations is the following orthogonality relation.

Proposition 10 (Orthogonality formula)**.** *Let $\{\, \chi_\alpha \,\}_\alpha$ be the set of characters of all irreducible representations of a group G. Then*

$$\frac{1}{|G|} \sum_{g \in G} (\chi_\alpha(g))^* \chi_\beta(g) = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{if } \alpha \neq \beta \end{cases}. \tag{17}$$

From this, follows one of the key results in representation theory is the formula for multiplicities, used to decompose a representation into irreducible ones.

Proposition 11 (Multiplicity formula)**.** *Let χ_α be the character of the irreducible representation π_α and ϕ the character of the representation π of a group G. Then*

$$\frac{1}{|G|}\sum_{g\in G}(\chi_\alpha(g))^*\phi(g) = m_\alpha, \tag{18}$$

where m_j is the multiplicity of the irreducible representation π_j in the decomposition of π, so that π is similar to a block diagonal matrix in the form

$$\pi(g) \simeq \bigoplus \pi_\alpha(g)\otimes \mathbb{1}_{m_\alpha} \qquad \forall g\in G, \tag{19}$$

with $\mathbb{1}_{m_j}$ being the identity matrix on $\mathbb{C}^{m_j}$.

For each irreducible subspace of V we can choose a basis set $\{\, v_j^\alpha \,\}_j$ where α is a label of the irreducible representation (with dimension $\dim\alpha$) decomposing π and $j \in \{\, 1,\ldots,\dim\alpha \,\}$. Each vector v in V can then be written as their linear combination $v = \sum_\alpha \sum_{j=1}^{\dim\alpha} c_j^\alpha\, v_j^\alpha$. We can conversely identify the basis vector components of any vector v by application of an appropriate projector P_j^α, such that $\mathrm{P}_j^\alpha\, v = c_j^\alpha\, v_j^\alpha$, where

$$\mathrm{P}_j^\alpha = \frac{\dim\alpha}{|G|}\sum_{g\in G}(\pi_\alpha(g))^*_{j,j}\,\pi(g). \tag{20}$$

Note that, in order to construct these projectors, the knowledge of the sole diagonal elements of the corresponding irreducible representation π_α is sufficient. By having access to the character table only, it is still possible to project any vector onto distinct irreducible subspaces (up to multiplicity) by using

$$\mathrm{P}^\alpha = \frac{\dim\alpha}{|G|}\sum_{g\in G}(\chi_\alpha(g))^*\,\pi(g). \tag{21}$$

C. Schur's Lemma

Hereby we write one of the most important results in representation theory, namely, *Schur's Lemma*. We will restrict it on finite-dimensional representations case.

Lemma 12 (Schur's Lemma)**.** *Let π_α and π_β be two irreducible representations of a finite group G of dimension m and n respectively, and M an $m\times n$ matrix. If*

$$\pi_\alpha(g)\, M\, \pi_\beta^{-1}(g) = M \qquad \forall g\in G \tag{22}$$

then π_α and π_β are equivalent irreducible representations or $M=0$.

Furthermore, if

$$\pi_\alpha(g)\, M\, \pi_\alpha^{-1}(g) = M \qquad \forall g\in G \tag{23}$$

then $M=\mu\mathbb{1}$, i.e., it is a scalar matrix.

D. Construction of irreducible representations of semi-direct product groups

It is possible to couple two groups to construct a new one using direct and semi-direct products. One can also obtain all their irreducible representations using knowledge about those of the original two groups alone. For direct products, the procedure is straightforward.

Theorem 13 ([44], Theorem 10, Chapter 3)**.** *Each irreducible representation of a direct group $G_1\times G_2$ is isomorphic to a representation $\pi_1\otimes\pi_2$ with π_1 and π_2 being irreducible representations of group G_1 and G_2 respectively*

For a group generated by a semi-direct product $N\rtimes H$, a more sophisticated machinery is needed (cfr. Refs. [44, 45]), and works only if the normal subgroup N is also *abelian*, i.e., all elements commute with respect to the group operation. Assuming N to be abelian, its irreducible representations $\{\, \chi_\alpha \,\}_\alpha$ are 1-dimensional and carry an action of G by

$$g\cdot\chi_\alpha(a) = \chi_\alpha(g^{-1}ag) \qquad \forall a\in N \text{ and } g\in G. \tag{24}$$

Now, consider the orbits of the characters induced by the action of H and choose a set of representatives $\{\, \chi_r \,\}_r$. For each r, let H_r be the stabilizer subgroup of χ_r in H and then define $G_r = N\cdot H_r$. Now extend χ_r to G_r by

$$\chi_r(ah) = \chi(a) \qquad \forall a\in N \text{ and } h\in H_r. \tag{25}$$

Let θ be an irreducible representation of H_r and lift it to an irreducible representation $\widetilde{\theta}$ of G_r through the canonical projection $P: G_r\to G_r/N$. As a final step, compose the two representations and obtain a representation $\rho_{r,\widetilde{\theta}}$ of the group G by *induction*, i.e., $\rho_{r,\widetilde{\theta}} = \mathrm{Ind}_{G_r}^G(\chi_r\cdot\widetilde{\theta})$. From Ref. [44, Proposition 25], we know that the so constructed representations $\rho_{r,\widetilde{\theta}}$ are irreducible and exhaust all irreducible representations of G. Since we will only need the characters $\chi_{\rho_{r,\widetilde{\theta}}}$ of the irreducible representations of G to apply Schur's Lemma, we will not elaborate what induced representations are. To obtain the sought characters, it suffices to make use of a Mackey-type formula,

$$\chi_{\rho_{r,\widetilde{\theta}}}(s) = \frac{1}{|G_r|}\sum_{\substack{g\in G\\ g^{-1}sg\in G_r}} \chi_r\cdot\chi_{\widetilde{\theta}}\,(g^{-1}sg). \tag{26}$$

E. Pauli-Liouville representation

To represent density operators and quantum channel on n qubits as vectors and matrices respectively, we will make use of the *Pauli-Liouville representation* with respect to Pauli basis. Let us pick

$$B \coloneqq \left\{ \frac{1}{\sqrt{2^n}}\bigotimes_{\nu=1}^{n}\tilde{\sigma}_\nu : \tilde{\sigma}_\nu \in \{\, \mathbb{1}_2, X, Y, Z \,\} \right\}, \tag{27}$$

where X, Y, Z are the Pauli matrices, as an orthonormal basis of $\mathrm{GL}(\mathbb{C}^{2n})$ with respect to the Hilbert-Schmidt inner product

$\langle A, B\rangle := \mathrm{Tr}[A^\dagger B]$. Then we know that we can express any density operator ρ and quantum channel $\mathcal{C}$ respectively as

$$\rho = \sum_{j=1}^{4^n} \rho_j \sigma_j \qquad \text{and} \qquad \mathcal{C}(\rho) = \sum_{j=1}^{4^n} \mathcal{C}(\sigma_j)\rho_j, \tag{28}$$

where $\sigma_j \in B$ and $\rho_j := \langle \sigma_j, \rho\rangle$, so that we can represent them as

$$|\rho\rangle = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_{4^n} \end{pmatrix} \qquad \text{and} \qquad \mathcal{C}_{jk} = \langle \sigma_j, \mathcal{C}(\sigma_k)\rangle. \tag{29}$$

In this way, we may transpose $\mathcal{C}(\rho)$ into a matrix-vector multiplication $\mathcal{C}\,|\rho\rangle$ and the concatenation of two channels $\mathcal{D}$ and $\mathcal{C}$ into a matrix multiplication $\mathcal{D}\,\mathcal{C}$. Additionally, representing channels in matrix form will allow us to make use of Schur's Lemma for matrix representations. We can analogously write a POVM E in the form

$$\langle E| = (E_1\ E_2 \dots E_{4^n}) \qquad \text{with} \qquad E_j = \langle E, \sigma_j\rangle. \tag{30}$$

With this, the probability to obtain an outcome described by E when measuring ρ is $p(E|\rho) = \mathrm{Tr}[E\rho] = \langle E, \rho\rangle$.

F. Zassenhaus formula

In order to justify our mathematical assumption through physical motivations, let us consider that the gate $\widetilde{U}$, which is the physical realization of the ideal gate U, is obtained during the application of some Hamiltonian H, perturbed for a fraction of time Δt (we denote the perturbed Hamiltonian as R), i.e., $\widetilde{U} = \mathrm{e}^{-\mathrm{i}(R\Delta t + HT)}$. Using the *Zassenhaus formula* [46], we can rewrite the implemented gate as

$$\widetilde{U} = \mathrm{e}^{-\mathrm{i}HT}\mathrm{e}^{-\mathrm{i}R\Delta t} \prod_{n=2}^{\infty} e^{C_n(HT, R\Delta t)} \tag{31}$$

$$= \mathrm{e}^{-\mathrm{i}HT}\left(\mathbb{1} - \mathrm{i}R\Delta t + \S\Delta t\right) + O(\Delta t^2), \tag{32}$$

where

$$\S := \sum_{n=2}^{\infty} c_n [\underbrace{H, [H, \dots, [H}_{n-1 \text{ times}}, R] \dots]] T^{n-1}, \tag{33}$$

with the Zassenhaus coefficients c_n that can be recursively calculated for instance as in Ref. [46]. This implies that the off-diagonal elements of the matrix $\widetilde{U}$ – computed in the eigenbasis of U – are of order Δt, justifying our assumption on the noise Λ.

G. First-order fitting model

We will now take into account off-diagonal matrix entries. A feature of the protocol being of great relevance is that, by twirling the error channel over the symmetry group G, we reduce the number of these off-diagonal matrix entries and so we drastically decrease the amount of parameters in the fitting model. Like in the main part of the letter, we write $\Lambda^G = \Lambda_0 + \Lambda_{\mathrm{off}}$, with Λ_0 being jointly diagonalizable with $\mathcal{U}$. Provided $\Lambda_{\mathrm{off}} = \{\mu_{i,j}\}_{i\neq j}$ to be "small" (i.e., the second order perturbation being negligible), we can consider the first-order model

$$F^{(1)}_{\mathrm{avg}}(\ell, \rho) = F^{(0)}_{\mathrm{avg}}(\ell, \rho) + \sum_{i\neq j}\sum_{p=0}^{\ell-1} (\lambda_i d_i)^p (\lambda_j d_j)^{\ell-p-1} \zeta_{i,j}, \tag{34}$$

with $\zeta_{i,j} := \mu_{i,j} d_j\ \mathrm{Tr}[E\,\Lambda'(v_i)]\langle \rho, v_j\rangle$ and the indices i, j labeling the elements within the support of ρ. This expression may be re-formulated into a simpler form, e.g., using the geometric series formula we obtain

$$F^{(1)}_{\mathrm{avg}}(\ell, \rho) = F^{(0)}_{\mathrm{avg}}(\ell, \rho) + \sum_{i\neq j} \frac{(\lambda_j d_j)^\ell - (\lambda_i d_i)^\ell}{\lambda_j d_j - \lambda_i d_i} \zeta_{i,j}. \tag{35}$$

As already mentioned, since we twirled over the symmetry group and so Λ^G is block-diagonal, a number of $\mu_{i,j}$ (and hence the corresponding $\zeta_{i,j}$) can be set to 0 in advance. More precisely, when a representation of the symmetry group is written as a direct sum of irreducible representations as

$$\pi(g) = \bigoplus_{\alpha\,\mathrm{irrep}} \mathbb{1}_{m_\alpha} \otimes \pi_\alpha(g), \tag{36}$$

where m_α is the multiplicity of the irreducible representation π_α, two matrices X and Y which are both commuting with $\pi(g)$ assume the form

$$X = \bigoplus_\alpha x^\alpha \otimes \mathbb{1}_{\dim \alpha} \quad \text{and} \quad Y = \bigoplus_\alpha y^\alpha \otimes \mathbb{1}_{\dim \alpha}, \tag{37}$$

where x^α, y^α are square matrices with $\dim x^\alpha = \dim y^\alpha = m_\alpha$. If X is normal, one can then choose a basis such that all x^α are diagonal (so that X will assume a diagonal form), while Y will maintain a similar form $Y = \bigoplus_\alpha \widetilde{y}^\alpha \otimes \mathbb{1}_{\dim \alpha}$. Hence, in our case, while diagonalizing $\mathcal{U}$ (from the Pauli-Liouville representation), Λ^G maintains a form as in Eq. (37).

H. Connecting to the average gate fidelity

For a quantum channel $\mathcal{E}$ and a unitary operation $\mathcal{U}$, the *gate fidelity* between these two quantities for a pure state ϕ is given by

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(\phi) := \mathrm{Tr}\left[\mathcal{U}(|\phi\rangle\langle\phi|)\,\mathcal{E}(|\phi\rangle\langle\phi|)\right] \tag{38}$$

and defining $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$ one has

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(\phi) = \mathcal{F}_{\Lambda,\mathcal{I}}(\phi) = \mathrm{Tr}\left[|\phi\rangle\langle\phi|\ \Lambda(|\phi\rangle\langle\phi|)\right], \tag{39}$$

hence quantifying the noise channel Λ for the implementation $\mathcal{E}$ of $\mathcal{U}$.

The *average gate fidelity* is then obtained by integrating this quantity over the Haar measure on pure states, that is,

$$\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{U}}) = \mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) \coloneqq \int_{\text{Haar}} \text{Tr}\left[|\phi\rangle\langle\phi|\ \Lambda(|\phi\rangle\langle\phi|)\right]\ \mathrm{d}\phi. \tag{40}$$

Conversely, the *entanglement fidelity* of a quantum channel $\mathcal{E}$, defined as

$$F_{\text{ent}}(\mathcal{E}) \coloneqq \langle\psi|\,(\mathcal{I}\otimes\ \mathcal{E})(|\psi\rangle\langle\psi|)\,|\psi\rangle\,, \tag{41}$$

with $|\psi\rangle$ being a maximally entangled state vector, can be written as [47]

$$F_{\text{ent}}(\mathcal{E}) = d^{-3}\sum_j \text{Tr}[V_j^\dagger \mathcal{E}(V_j)], \tag{42}$$

for any orthonormal basis $\{\,V_j\,\}_j$ such that $\text{Tr}[V_jV_k] = d\,\delta_{j,k}$ (in the case of n qubits, $d = 2^n$). The average gate fidelity of $\mathcal{E}$ is then linked to this quantity by [47]

$$\mathbb{E}(\mathcal{F}_{\mathcal{E},\mathcal{I}}) = \frac{dF_{\text{ent}}(\mathcal{E})+1}{d+1} = \frac{\sum_j \text{Tr}[V_j^\dagger\mathcal{E}(V_j)]+d^2}{d^2(d+1)}, \tag{43}$$

so that the average gate fidelity of the twirled error channel Λ^G is related to the parameters $\{\,\lambda_j\,\}_j$ obtained in the fitting model in Eq. (7) from the main part of the letter by

$$\mathbb{E}(\mathcal{F}_{\Lambda^G,\mathcal{I}}) = \frac{\sum\lambda_j + d}{d(d+1)}. \tag{44}$$

Now the question is what information about the original noise channel we can extract from the twirled channeld Λ^G. In fact they are the same, since the entanglement fidelity is invariant under twirling over the symmetry group G. Let us rewrite

$$\begin{aligned} F_{\text{ent}}(\Lambda^G) &= d^{-3}\sum_j \text{Tr}[V_j^\dagger\Lambda^G(V_j)] && (45)\\ &= \frac{d^{-3}}{|G|}\sum_{k=1}^{|G|}\sum_j \text{Tr}[V_j^\dagger g_k^\dagger\Lambda(g_kV_jg_k^\dagger)g_k] && (46)\\ &= \frac{d^{-3}}{|G|}\sum_{k=1}^{|G|}\sum_j \text{Tr}[(W_j^k)^\dagger\Lambda(W_j^k)], && (47)\end{aligned}$$

where we denote $W_j^k = g_kV_jg_k^\dagger$ and used cyclicity of the trace. Since W_j^k is again an orthogonal basis with respect to the Hilbert-Schmidt inner product (i.e., $\text{Tr}[(W_{j'}^k)^\dagger W_j^k] = d\,\delta_{j'j}\ \forall k$), then $d^{-3}\sum_j \text{Tr}[(W_j^k)^\dagger\Lambda(W_j^k)] = F_{\text{ent}}(\Lambda)$ so that

$$F_{\text{ent}}(\Lambda^G) = \frac{1}{|G|}\sum_{k=1}^{|G|} F_{\text{ent}}(\Lambda) = F_{\text{ent}}(\Lambda) \tag{48}$$

and hence

$$\mathbb{E}(\mathcal{F}_{\Lambda,\mathcal{I}}) = \mathbb{E}(\mathcal{F}_{\Lambda^G,\mathcal{I}}). \tag{49}$$

I. Characterizing the error of the single gate U

In order to recover the fidelity of the gate U from the noise Λ, which originates from the composition of U and a unitary gate from the symmetry group G, we first consider the χ *matrix representation of* $\mathcal{E}$,

$$\mathcal{E}(\rho) = \sum_{i,j}\chi_{i,j}P_i\rho P_j\ , \tag{50}$$

where $P_0 = \mathbb{1}$. We can characterize the error of the gate U, distinguishing it from the one coming from the symmetry group G, that we consider to be $\mathcal{N}$ for all elements in the group (which can be benchmarked separately using for instance the known methods to benchmark Clifford gates), using the bound from Ref. [48, Appendix D] (where we set $i = 0$)

$$\begin{aligned}|\chi_{0,0}^{\Lambda\circ\mathcal{N}} - \chi_{0,0}^{\Lambda}\chi_{0,0}^{\mathcal{N}}| \le 2\left((1-\chi_{0,0}^{\Lambda})\chi_{0,0}^{\Lambda}(1-\chi_{0,0}^{\mathcal{N}})\chi_{0,0}^{\mathcal{N}}\right)^{1/2}\\ + (1-\chi_{0,0}^{\Lambda})(1-\chi_{0,0}^{\mathcal{N}}).\end{aligned} \tag{51}$$

For an arbitrary channel $\mathcal{E}$, we know that $\chi_{0,0}^{\mathcal{E}} = \text{Tr}[\mathcal{E}]/d^2$ (cfr. [49, Eq. (2.30)] and Eq. (44)), so that we can recover the fidelity of the gate U form the ones of the gates belonging to G and from $\mathbb{E}(\mathcal{F}_{(\Lambda\circ\mathcal{N})^G,\mathcal{I}})$ obtained with our protocol. The bound in particular is valid in the regime $\chi_{0,0}^{\mathcal{N}} \approx 1$, i.e., when the gates of the symmetry group can be implemented with high fidelity.

J. Confidence interval

This section is concerned with uncertainty quantification in our scheme. To assess the number of different random sequences that have to be sampled in order to justify $F_{\text{seq}}(\ell) \approx F_{\text{avg}}(\ell)$ for a given sequence length ℓ, Wallman and Flammia in Ref. [50] provided bounds on the variance for the Clifford randomized benchmarking protocol described in Ref. [49]. Their results show that a relatively small number of random samples is needed. We want to show a bound similar to that of Ref. [50, Theorem 10],

$$\sigma_\ell^2 = \frac{1}{|G|^\ell}\sum_{\mathbf{k}_\ell} F_{\mathbf{k}}^2(\ell,\rho) - F_{\text{avg}}(\ell,\rho)^2. \tag{52}$$

In Pauli-Liouville representation and using $(E|\mathcal{C}|\rho)^2 = (E^{\otimes 2}|\mathcal{C}^{\otimes 2}|\rho^{\otimes 2})$, this can be expressed in terms of a scalar product in the form

$$\sigma_\ell^2 = \frac{1}{|G|^\ell}\sum_{\mathbf{k}_\ell}(E^{\otimes 2}|\mathcal{C}_{\mathbf{k}_\ell}^{\otimes 2}|\rho^{\otimes 2}) - (E^{\otimes 2}|\mathcal{C}_{\text{avg},\ell}^{\otimes 2}|\rho^{\otimes 2}). \tag{53}$$

Now, we assume to be in the regime $\Lambda = \mathbb{1}+Q\Delta t$, where Q is a bounded matrix under additional assumption $\text{Tr}Q = \Theta(d^2)$, and expand the expression for the variance up the second order in Δt.

$$\sigma_\ell^2 = \Delta t^2 (E^{\otimes 2}|\sum_{j=1}^{\ell} \frac{1}{|G|} \sum_{B\in G} (\mathcal{U}\otimes\mathcal{U})^{\ell-j}(\mathcal{B}^\dagger Q \mathcal{B}\otimes \mathcal{B}^\dagger Q\mathcal{B})(\mathcal{U}\otimes\mathcal{U})^j - \sum_{j=1}^{\ell}(\mathcal{U}\otimes\mathcal{U})^{\ell-j}(Q^G\otimes Q^G)(\mathcal{U}\otimes\mathcal{U})^j|\rho^{\otimes 2}) + O(\ell^2 r^2 d^4). \tag{54}$$

The first term can be bounded as in Ref. [50] using diamond norm properties and Ref. [50, Proposition 9] with $4d(d+1)\ell r$. Again following that argument, the terms of order $O(\Delta t^3 Q^3)$ and $O(\Delta t^4 Q^4)$ are $O(\ell^2 r^2 d^4)$. Knowing the structure of Q^G from the analysis of the symmetry group G, we can upper bound the number of non-zero terms as

$$\sum_\alpha m_\alpha^2\, d_\alpha \leq \max{}_\alpha m_\alpha \sum_\alpha m_\alpha\, d_\alpha = \max{}_\alpha m_\alpha\, d^2. \tag{55}$$

From now on, we denote $m = \max{}_\alpha m_\alpha$ and $q = \max{}_{i,j} q_{i,j}$, the largest matrix entry of Q that we assume being independent of d. The second term in expression (54) obeys to the inequality

$$(E^{\otimes 2}|\sum_{j=1}^{\ell}(\mathcal{U}\otimes\mathcal{U})^{\ell-j}(Q^G\otimes Q^G)(\mathcal{U}\otimes\mathcal{U})^j|\rho^{\otimes 2}) \leq \ell q^2 m^2 d^4. \tag{56}$$

Using

$$\mathrm{Tr}[\Lambda] = d(d+1)\mathbb{E}(\mathcal{F}) - d \tag{57}$$

follows

$$\Delta t = -r\frac{d(d+1)}{\mathrm{Tr}[Q]}, \tag{58}$$

and so $\Delta t = O(r)$ since we assumed $\mathrm{Tr}Q = \Theta(d^2)$. Hence, second term of Eq. (54) is $O(m^2\,\ell\, r^2 d^4)$. Summarizing gives a bound for the variance as

$$\sigma_\ell^2 \leq 4d(d+1)\ell r + O(\ell^2 r^2 d^4) + O(m^2\,\ell\, r^2 d^4), \tag{59}$$

where the second term dominates the third one for $\ell \gg m^2$, i.e., in this regime the bound is exactly equivalent to the one of Ref. [50, Theorem 10]. This bound however is probably not tight, and we are interested whether a bound similar to the one provided in Ref. [51] can be obtained.

K. Example of tensor copies of T-gates

Starting from this part of the appendix we will illustrate the application of our scheme to tensor copies of the T-gate, one of the most relevant quantum gates, which together with the H, S and CNOT gates gives rise to a universal quantum circuit. Specifically within the context of fault tolerant quantum computing, this situation is of paramount importance. First, we are going to depict the irreducible decomposition up to four tensor copies of this gate in order to get a feeling on the scalability and necessary resources for this method. We give in the following the step-by-step sequence according to the procedure described in Appendix D.

[1] Produce the n-Kronecker product group, denoted by A_n, of the local abelian symmetry group

$$\left\{ \begin{pmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1 \end{pmatrix}, \begin{pmatrix} 1&0&0&0\\0&0&-1&0\\0&1&0&0\\0&0&0&1 \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} 1&0&0&0\\0&0&1&0\\0&-1&0&0\\0&0&0&1 \end{pmatrix}, \begin{pmatrix} 1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&1 \end{pmatrix} \right\},$$

which is isomorphic to the *cyclic group* of order 4, $\mathbb{Z}_4$.

[2] Construct representation of the symmetric group S_n permuting the local subsystems.

[3] Construct the full symmetry group G as a semi-direct product of A_n and S_n by multiplying the respective 4^n-dimensional matrix representations. Each $g \in G$ is given by $g = a.\sigma$, with $a \in A_n, \sigma \in S_n$.

[4] From the character table of $\mathbb{Z}_4$,

$\mathbb{Z}_4$	e	γ	γ^2	γ^3
χ_0	1	1	1	1
χ_1	1	i	-1	-i
χ_2	1	-1	1	-1
χ_3	1	-i	-1	i

construct the character table of A_n by taking the product of the respective characters

$$\chi_{c_1,c_2,\ldots,c_n}(\ell_1,\ell_2,\ldots,\ell_n) \coloneqq \chi_{c_1}(\ell_1)\chi_{c_2}(\ell_2)\cdots\chi_{c_n}(\ell_n)\,, \tag{60}$$

where $\ell_j \in \mathbb{Z}_4$ and c_j is the label representing the irreducible representation.

[5] Divide the characters of A_n into orbits with respect to the action of S_n given by

$$\sigma.\chi(a)_{c_1,c_2,\ldots,c_n} \coloneqq \chi(\sigma^{-1}a\,\sigma)_{c_1,c_2,\ldots,c_n}. \tag{61}$$

In this particular case, the action of S_n works as a permutation of the labels of the irreducible representations, i.e.,

$$\sigma.\chi(a)_{c_1,c_2,\ldots,c_n} = \chi(a)_{\sigma(c_1,c_2,\ldots,c_n)}. \tag{62}$$

Choose for each orbit a representative element, for instance $\chi(a)_{c_1,c_2,\ldots,c_n}$ with $c_1 \leq c_2 \leq \cdots \leq c_n$, building a set $\{\,\chi_j\,\}_j$.

[6] For each representative element χ_j, find the stabilizer group H_j as a subgroup of S_n.

[7] For each irreducible representation π of H_j, write an irreducible representation of the subgroup $G_j \coloneqq A_n \cdot H_j$ of G by

$$\tilde{\rho}^j_\pi(a, g_j) = \chi_j(a) \cdot \pi(g_j). \tag{63}$$

[8] Obtain the characters of the representation ρ^j_π of G induced by $\tilde{\rho}^j_\pi$ with the Mackey-type formula,

$$\chi_{\rho^j_\pi}(s) = \frac{1}{|G_j|} \sum_{\substack{t\in G \\ t^{-1}st\in G_j}} \chi_{\tilde{\rho}^j_\pi}(t^{-1}st), \tag{64}$$

and obtain the irreducible representation multiplicity m^j_π in the decomposition of the Pauli-Liouville representation of G with the formula

$$m^j_\pi = \frac{1}{|G|} \sum_{g\in G} \left(\chi_{\rho^j_\pi}(g)\right)^* \cdot \phi(g), \tag{65}$$

where $g \mapsto \phi(g)$ is the trace of g in Pauli-Liouville representation.

In case of $n = 4$, for instance, there are 256 different irreducible representations of A_4 and five stabilizer groups: the full permutation group S_4 for the irreducible representations of the form $\chi_{a,a,a,a},\ a \in 0, 1, 2, 3$, giving rise to $4 \cdot 5 = 20$ irreducible representations for G, S_3 for the representative irreducible representations of the form $\chi_{a,a,a,b}$ and $\chi_{a,b,b,b}$ with $a < b$, giving rise to $12 \cdot 3 = 36$ new irreducible representations, $S2 \times S2$ (isomorphic to the Klein four-group) for representative elements $\chi_{a,a,b,b}$ with $a < b$, so that a total of $6 \cdot 4 = 24$ irreducible representations of G are derived, again representative elements $\chi_{a,a,b,c}, \chi_{a,b,b,c}, \chi_{a,b,c,c}$ with $a < b < c$ have stabilizer S_2, producing additional $12 \cdot 2 = 24$ irreducible representations; finally, the single element $\chi_{0,1,2,3}$ represents the sole orbit with trivial stabilizer. Hence, we have in total 105 different induced irreducible representations of G whose characters are obtained using Eq. (64). As one can observe in Table IV, only 22 of these irreducible representations decompose the twirled noise matrix, and the trivial representation has the highest multiplicity.

Results for $n \leq 4$

We report in Tables I-IV the irreducible decompositions for up to four tensor copies of the T-gate. The superscripts of χ label the irreducible representations of A_n, while after the semicolon we denote the irreducible representation of the stabilizer group, where e denotes the trivial representation, sgn the sign representation, std the standard representation for all subgroups of S_4, $\ker_a$ the *Kernel a* representation of the Klein four-group isomorphic to $S_2 \times S_2$, while $2\mathrm{dim}$ denotes the 2-dimensional representation of S_4. We note that χ_2 never appears in the decomposition, and that the highest multiplicity, being $n + 1$, is always related to the trivial representation of the full group G. Additionally, we note that exploiting Schur's Lemma and the above consideration, the number of distinct λ_j to be fitted when benchmarking copies of the T-gate is $\sum_{\alpha\ \mathrm{irrep}} m_\alpha$, i.e., from 1 to 4 qubits, this number is $4, 11, 24, 46$.

L. Classical recovery and estimation methods

In this section, we give details on the our numerical methods employed to recover the average fidelity from the measurement data provided by our protocol. In principle, we could try to obtain the parameters directly by a non-linear fitting approach along the lines of the variable projection algorithm that separates fitting of linear and non-linear parameters [52]. However, given the fact that the quantities we want to obtain correspond to the estimation of different decay rates in the data set, we are going to use instead a matrix pencil method for the extraction of signal poles developed in the context of *signal processing* [53, 54]. Due to the random nature of the sampling paths, this is supplemented by a bootstrapping approach in order to get reliable estimates on these parameters. We will continue describing these two components of our methodology in detail now, before commenting on the single and two T-gate examples described in the main text.

1. *Matrix pencil methods*

Key to the functioning of our scheme is the use of sophisticated methods of estimation which we lay out here. According to (44), we can express the average fidelity $\mathbb{E}(\mathcal{F}_{\Lambda^G,\mathcal{I}})$ of the twirled channel in terms of the model parameters λ_j as

$$\mathbb{E}(\mathcal{F}_{\Lambda^G,\mathcal{I}}) = \frac{\sum \lambda_j + d}{d(d+1)}. \tag{66}$$

The relation between the model parameters λ_j and the measurement data on the other hand is given by the zeroth-order fitting model

$$F^{(0)}_{\mathrm{avg}}(\ell, \rho) = \sum_{j=1}^{4^n} (\lambda_j\, d_j)^\ell \xi_j\ . \tag{67}$$

From an abstract point of view this means that up to higher order terms, for each $\ell = 1, \ldots, \ell_{\max}$, the measurement result F_ℓ can be expressed as a sum exponentially decaying terms of the form

$$F_\ell = \sum_j^M \xi_j x_j^\ell \tag{68}$$

where we set $x_j = \lambda_j d_j$. Recovering the parameters x_j from such a noisy data set is a well studied problem in the context of signal reconstruction going back to the work by Prony. Modern algorithms known as the *Estimation of signal parameters*

Irreducible representation	Multip.
χ_0	2
χ_1	1
χ_3	1

Table I. This and the subsequent three tables depict the irreducible decompositions of the symmetry group G of multiple tensor copies of the T-gate channel. The superscripts of χ label the irreducible representations of A_n, the word after the semicolon the irreducible representation of the stabilizer group: e denotes the trivial representation, sgn the sign representation, std the standard representation, $\ker_a$ the *Kernel* a representation of the Klein four-group, $2\mathrm{dim}$ the 2-dimensional representation of S_4.

Irreducible representation	Dim	Multip.	Irreducible representation	Dim	Multip.
$\chi_0, \chi_0; e$	1	3	$\chi_0, \chi_1; e$	2	2
$\chi_1, \chi_1; e$	1	1	$\chi_0, \chi_3; e$	2	2
$\chi_3, \chi_3; e$	1	1	$\chi_1, \chi_3; e$	2	1
$\chi_0, \chi_0, \mathrm{sgn}$	1	1			

Table II. Two T-gates decomposition.

via rotational Invariance Techniques (ESPRIT) rely on feature extraction via singular value decompositions and matrix pencils [55]. For a recent review and details of the construction see, e.g., Ref. [56]. In the following, we give a short overview over the algorithm.

In order to extract at most $M_{\max}$ signal poles x_i from $\{F(\ell)\}_{\ell=0}^{L_{\max}-1}$, we first form the $2L_{\max} - M_{\max} \times M_{\max} + 1$ *Hankel matrix* $H_{\ell,k} = F(\ell + k)$ and compute its singular value decomposition

$$H = UDV^*. \tag{69}$$

If there were no noise in the data, the number of non-zero singular values of H would correspond directly to the number of unique $\{x_i\}$ in (68). However, in the presence of noise, H will typically feature full rank and we have to fix a threshold $\sigma_{\min}$ for the singular values to obtain a rank M approximation of H in order to extract $M \leq M_{\max}$ poles for the approximation of our measurement data. Let us denote by W the $M \times M_{\max} + 1$ matrix formed by the first M rows of V^* and set the $M_{\max} \times M$ matrices V_0 and V_1 equal to $(V_i)_{\ell,k} = M_{k,\ell+i}$, $\ell = 1, \ldots, M_{\max}$, $k = 1, \ldots, M$. The estimate for the signal poles x_i can now be computed as the M eigenvalues of the matrix $(V_0)^{-1} \cdot V_1$, where the inverse of V_0 is defined as its Moore-Penrose pseudo-inverse. In this way, we can extract from (67) the parameters $\{x_j\} = \{\lambda_j d_j\}$.

Now in principle, we would be left with the problem of matching the right x_j with the correct d_j in order to obtain the parameters λ_j. However, since typical quantum gates and in particular the T-gate have eigenvalues which are close to being roots of unity, we can exploit this property to circumvent this problem. Namely, assuming that we can find an τ such that $d_j^\tau = 1$ for all j, we can partition our measurement data $\{F(\ell)\}_{\ell=1}^{\ell_{\max}}$ into τ subsets of the form $\{F(r), F(r+\tau), F(r+2\tau), \ldots\}$ with $r = 1, \ldots, \tau$. In each of these subsets (67) reads as

$$F_r(\ell) = \sum_j \xi_j d_j^r \left(d_j^\tau\right)^\ell \left(\lambda_j^\tau\right)^\ell = \sum_j \widetilde{\xi}_{j,r} \left(\lambda_j^\tau\right)^\ell \ , \tag{70}$$

which is again of the form (68) with $x_j = \lambda_j^\tau$ and $\widetilde{\xi}_{j,r} = \xi_j d_j^r$. Now using the reconstruction method described before, we can extract from this data λ_j^τ and in turn obtain d_j by taking the τs-root. In terms of data collection from an experiment, we now have two options: we can restrict the set of data points, i.e., measured expectation values, to $\ell = 1, \tau, 2\tau, \ldots$ or we can also collect data at intermediate points and extract the poles from the combined Hankel matrix $[H_1, H_2, \ldots, H_\tau]$ in the way described before according to a multi-channel signal reconstruction approach [57].

2. *Restriction to symmetry subspaces*

In this subsection, we will explain how the restriction to symmetry subspaces will lead to a robust numerical recovery procedure. Under our standing assumption that the twirled noise channel Λ_U^G is almost jointly diagonalizable with the target unitary channel $\mathcal{U}$, it follows directly that Λ_U^G will also approximately preserve the invariant subspaces of the symmetry group G. For our protocol this means that by choosing an appropriate initial states ρ supported on such an invariant subspace, $(\Lambda_U^G)^\ell(\rho)$ will still be approximately located there. This observation serves us in two ways: On the one hand, we can get additional information about the performance of the benchmarked operation with respect to a specific subspace, i.e., we can identify the subspaces on which the errors occur. On the other hand, we reduce the number of signal parameters λ_j we have to extract from our data considerably if we restrict the initial state to a particular subspace. This becomes particularly important when we come close to an optimal implementation where the λ_j become closer and closer to being degenerate. To apply this approach, previously developed to tackle multi-parameters fitting models for finite groups [57], we first construct the seven projectors linked to the irreducible representations in Table II and then take the corresponding eigenvectors as a basis of the invariant subspaces. Subsequently, we need to construct from these vectors a set of density operators such that we can address each irreducible subspace at

Irreducible representation	Dim	Multip.	Irreducible representation	Dim	Multip.	Irreducible representation	Dim	Multip.
$\chi_0, \chi_0, \chi_0; e$	1	4	$\chi_0, \chi_0, \chi_3; e$	3	3	$\chi_0, \chi_0, \chi_1; \mathrm{sgn}$	3	1
$\chi_1, \chi_1, \chi_1; e$	1	1	$\chi_0, \chi_1, \chi_1; e$	3	2	$\chi_0, \chi_0, \chi_3; \mathrm{sgn}$	3	1
$\chi_3, \chi_3, \chi_3; e$	1	1	$\chi_1, \chi_1, \chi_3; e$	3	1	$\chi_0, \chi_1, \chi_3; e$	6	2
$\chi_0, \chi_0, \chi_0;$ std	2	2	$\chi_0, \chi_3, \chi_3; e$	3	2			
$\chi_0, \chi_0, \chi_1; e$	3	3	$\chi_1, \chi_3, \chi_3; e$	3	1			

Table III. Three T-gates decomposition.

Irreducible representation	Dim	Multip.	Irreducible representation	Dim	Multip.	Irreducible representation	Dim	Multip.
$\chi_0, \chi_0, \chi_0, \chi_0; e$	1	5	$\chi_1, \chi_1, \chi_1, \chi_3; e$	4	1	$\chi_1, \chi_1, \chi_3, \chi_3; e$	6	1
$\chi_1, \chi_1, \chi_1, \chi_1; e$	1	1	$\chi_0, \chi_3\chi_3, \chi_3; e$	4	2	$\chi_0, \chi_0, \chi_1, \chi_1; \ker_a$	6	1
$\chi_3, \chi_3, \chi_3, \chi_3; e$	1	1	$\chi_1, \chi_3\chi_3, \chi_3; e$	4	1	$\chi_0, \chi_0, \chi_3, \chi_3; \ker_a$	6	1
$\chi_0, \chi_0, \chi_0, \chi_0; 2\mathrm{dim}$	2	1	$\chi_0, \chi_0, \chi_0, \chi_1; \mathrm{std}$	8	2	$\chi_0, \chi_0, \chi_1, \chi_3; e$	12	3
$\chi_0, \chi_0, \chi_0, \chi_0; \mathrm{std}$	3	3	$\chi_0, \chi_0, \chi_0, \chi_3; \mathrm{std}$	8	2	$\chi_0, \chi_1, \chi_1, \chi_3; e$	12	2
$\chi_0, \chi_0, \chi_0, \chi_1; e$	4	4	$\chi_0, \chi_0, \chi_1, \chi_1; e$	6	3	$\chi_0, \chi_1, \chi_3, \chi_3; e$	12	2
$\chi_0, \chi_0, \chi_0, \chi_3; e$	4	4	$\chi_0, \chi_0, \chi_3, \chi_3; e$	6	3	$\chi_0, \chi_0, \chi_1, \chi_3; \mathrm{sgn}$	12	1
$\chi_0, \chi_1, \chi_1, \chi_1; e$	4	2						

Table IV. Four T-gates decomposition.

least once and target only few of those subspaces in a single iteration. This choice is clearly not unique; for our numerics, we selected as input state the density matrices given in Section M. It is a key insight to the functioning of the method that we can address single or a few irreducible subspaces in each iteration to arrive at a reliable and robust method that is able to reliably discriminate between close poles.

3. *Bootstrapping for parameter estimation*

Here, we give further details on the our numerical methods used to extract the average fidelity from the measurement data provided by our protocol. The starting point is the observation that the data obtained experimentally by executing our scheme contain more information than necessary for the model: Instead of having access to the observed average survival probabilities for a given execution length ℓ solely, our protocol actually provides this information at the level of each of the randomly chosen sample paths of length ℓ individually. This insights motivates the use of standard methods in statistical estimation called *bootstrapping procedures* [59] in order to reliably extract the average fidelity. Bootstrapping techniques refer to random sampling methods with replacement, designed in order to assign measures of accuracy to sample estimates.

For this, note that the sampling paths both with respect to the sequence length ℓ as well as with respect to the given realization of length ℓ are chosen independently. Hence, given the set of measured survival probabilities $\{F_{\mathbf{k}_{\ell,q}}\}$, where $1 \leq \ell \leq \ell_{\max}$ denotes the circuit length and $\mathbf{k}_{\ell,q}$ the q-th randomly chosen sample sequence of symmetry gates for length ℓ, we are going to resample this set of data with respect to ℓ. More precisely, in order to create a single bootstrap sample, we pick for each fixed ℓ a random subset with replacement of m elements from $\{F_{\mathbf{k}_{\ell,q}}\}$ and compute their average with respect to q. For each of these resampled sequences, we compute the approximated average fidelity, according to our fitting model and matrix pencil methods described in Appendix L 1. The final estimate for the averaged fidelity is then obtained as average over all taken bootstrap samples.

M. Numerical recovery

In this section, we apply the recovery procedure detailed in Appendix L 1 for two significant examples of practical relevance that are problematic to handle with the common approach to randomized benchmarking: A single T-gate and the tensor product of two T-gates.

1. *Single T-gate-case*

We describe here the numerics reported in the main text in greater detail. First, we consider the case of a single T-gate that is generated by the noisy Hamiltonian

$$H_\varepsilon = \frac{\pi}{8}\sigma_z - \varepsilon\,\sigma_x \tag{71}$$

with noise parameter $\varepsilon > 0$. The four eigenvalues of the T-gate channel including multiplicities are given by $\{1, 1, (1+\mathrm{i})/\sqrt{2}, (1-\mathrm{i})/\sqrt{2}\}$. Accordingly, we have to extract four decay parameters λ_j to estimate the average fidelity from (67). As described in Appendix L 1, we make use of the fact that the $T^8 = \mathbb{1}$ and hence we can extract the parameters λ_j by described matrix pencil approach. By exploiting the symmetry of the T-gate, we can obtain information in three invariant subspaces by running our protocol with the initial

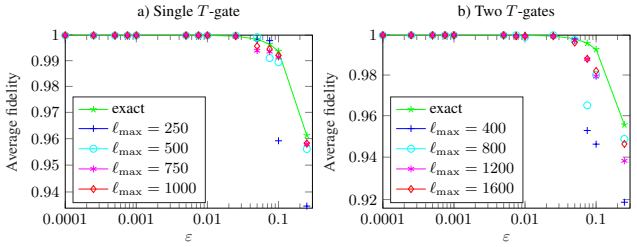

Figure 1. Dependence of the estimated average fidelity for noisy single and two T-gate on the maximal sequence length for different noise-strengths ε extracted from bootstrapping over $N = 100$ sequences of length the given length (b). Green lines and stars indicate the analytic value of the average fidelity $\mathbb{E}(\mathcal{F})$ for the given noise level.

states

$$\rho \in \left\{ \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & -\mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \right\} \tag{72}$$

and measuring the corresponding survival probability. Fig. 1 (a) shows the extracted estimates of the average fidelity depending on the strength of the perturbation ε and the maximal sequence length of the protocol. It can be seen that we achieve a good agreement with the analytical value of the average fidelity even for short protocol sequences in the case of perturbations smaller than $\varepsilon \leq 0.01$. However, by increasing the sequence length to $\ell_{\max} = 1000$, we even achieve satisfactory estimates starting from $\varepsilon \leq 0.1$.

2. *Two T-gate case*

In the case of two T-gates applied in parallel to two qubits, we consider the perturbed Hamiltonian

$$H_\varepsilon = \frac{\pi}{8}\left(\sigma_z \otimes \mathbb{1} + \mathbb{1} \otimes \sigma_z - \varepsilon\, \sigma_x \otimes \sigma_x\right) \tag{73}$$

with the additional noise-term $\sigma_x \otimes \sigma_x$. In order to extract the decay parameters λ_j, we follow the approach of the single T-gate, the main difference being that we now choose the initial state among eight density matrices supported on different symmetry subspaces (they are reported below this paragraph). In each of these subspaces, we extract the corresponding decay parameters λ_j, which after bootstrapping leads us to an estimate of the desired average fidelity. Fig. 1 (b) depicts these estimates both in relation to the perturbation strength ε as well as depending on the maximally protocol length. Similar to the single T-gate case, we see that $\varepsilon \leq 0.01$ already $\ell_{\max} = 800$ gives satisfactory results. However, in order to obtain meaningful lower bounds for larger values of ε, we have to double $\ell_{\max}$ to 1600.

We now provide the quantum states used as input for our numerics. Each of them combines the totally mixed state with some of the basis vectors obtained by the projectors onto the irreducible subspaces. To obtain the diagonal elements of Λ connected to the trivial subspaces, we choose the states

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix} \text{ and } \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

With the first one we can extract a single pole and with each of the other two states an additional parameter. With

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & -\frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ -\frac{1}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

we retrieve two new parameters (in addition to the one corresponding to the trivial representation linked to the totally mixed state), one for irrep $\chi_1, \chi_1; e$ and one for $\chi_3, \chi_3; e$. The fifth state,

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix},$$

is a combination of the totally mixed state and the one belonging to the irrep $\chi_0, \chi_0, \mathrm{sgn}$. With the first of the following two density matrices,

$$\begin{pmatrix} \frac{1}{4} & -\frac{\mathrm{i}}{8} & -\frac{\mathrm{i}}{8} & 0 \\ \frac{\mathrm{i}}{8} & \frac{1}{4} & 0 & \frac{\mathrm{i}}{8} \\ \frac{\mathrm{i}}{8} & 0 & \frac{1}{4} & \frac{\mathrm{i}}{8} \\ 0 & -\frac{\mathrm{i}}{8} & -\frac{\mathrm{i}}{8} & \frac{1}{4} \end{pmatrix}, \begin{pmatrix} \frac{1}{4} & -\frac{\mathrm{i}}{8} & -\frac{\mathrm{i}}{8} & 0 \\ \frac{\mathrm{i}}{8} & \frac{1}{4} & 0 & -\frac{\mathrm{i}}{8} \\ \frac{\mathrm{i}}{8} & 0 & \frac{1}{4} & -\frac{\mathrm{i}}{8} \\ 0 & \frac{\mathrm{i}}{8} & \frac{\mathrm{i}}{8} & \frac{1}{4} \end{pmatrix},$$

we obtain two poles for $\chi_0, \chi_1; e$ and two for $\chi_0, \chi_3; e$ (and again the trivial one). With the second operator with get the other two poles of these two irreducible representations. The last density matrix,

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{\frac{1}{4}-\frac{\mathrm{i}}{4}}{\sqrt{2}} & 0 \\ 0 & \frac{\frac{1}{4}+\frac{\mathrm{i}}{4}}{\sqrt{2}} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix},$$

contains the basis vectors of the 2-dim irrep $\chi_1, \chi_3; e$. With this selection, we cover every irreducible subspace and so we can extract all the λ_j elements, addressing at most five of them at one time.

[1] A. Acin, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm. The European quantum technologies roadmap. *New J. Phys.*, 20:080201, 2018.

[2] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.*, 44:2455–2467, 1997.

[3] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, 2008.

[4] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, 2010.

[5] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch. Recovering quantum gates from few average gate fidelities. *Phys. Rev. Lett.*, 121:170502, 2018.

[6] M. Kliesch, R. Kueng, J. Eisert, and D. Gross. Improving compressed sensing with the diamond norm. *IEEE Trans. Inf. Th.*, 62:7445, 2016.

[7] A. Steffens, M. Friesdorf, T. Langen, B. Rauer, T. Schweigler, R. Hübener, J. Schmiedmayer, C. A. Riofrio, and J. Eisert. Towards experimental quantum field tomography with ultracold atoms. *Nature Comm.*, 6:7663, 2015.

[8] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos. Efficient tomography of a quantum many-body system. arXiv:1612.08000.

[9] J. Preskill. Fault-tolerant quantum computation. *arXiv:quant-ph/9712048*, 1997.

[10] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508, 2014.

[11] J. M. Chow, J. M. Gambetta, E. Magesan, S. J. Srinivasan, A. W. Cross, D. W. Abraham, N. A. Masluk, B. R. Johnson, C. A. Ryan, and M. Steffen. Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nature Comm.*, 5:4015, 2014.

[12] J. Emerson, R. Alicki, and K. Zyczkowski. Scalable noise estimation with random unitary operators. *J. Opt. B*, 7:S347–S352, 2005.

[13] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *arXiv:0606161*, 2006.

[14] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, 2008.

[15] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, 2012.

[16] E. Magesan, J. M. Gambetta, and J. Emerson. Robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:042311, 2011.

[17] J. J. Wallman and S. T. Flammia. Randomized benchmarking with confidence. *New. J. Phys.*, 16:103032, 2014.

[18] K. Boone, A. Carignan-Dugas, J. J. Wallman, and J. Emerson. Randomized benchmarking under different gatesets. *arXiv:1811.01920*.

[19] D. Gross, K. Audenaert, and J. Eisert. Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.*, 48:052104, 2007.

[20] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta. Scalable randomised benchmarking of non-Clifford gates. *npj Quant. Inf.*, 2, 2016.

[21] C. A. Dugas, J. Wallman, and J. Emerson. Characterizing universal gate sets via dihedral benchmarking. *Phys. Rev. A*, 92:060302, 2015.

[22] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.*, 109:080505, 2012.

[23] R. Harper and S. T. Flammia. Estimating the fidelity of T gates using standard interleaved randomized benchmarking. *Quant. Sc. Tech.*, 2:015008, 2017.

[24] J. J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, 2018.

[25] D. S. Franca and A. K. Hashagen. Approximate randomized benchmarking for finite groups. *J. Phys. A*, 51:395302, 2018.

[26] W. G. Brown and B. Eastin. Randomized benchmarking with restricted gate sets. *Phys. Rev. A*, 97:062323, 2018.

[27] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38:814–824, May 1990.

[28] 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*, 4:011050, 2014.

[29] The supplements cite Refs. [44–57, 59].

[30] K. Glashoff and M. M. Bronstein. Almost-commuting matrices are almost jointly diagonalizable, 2013. arXiv:1305.2135.

[31] B. E. Anderson, H. Sosa-Martinez, C. A. Riofrio, I. H. Deutsch, and P. S. Jessen. Accurate and robust unitary transformations of a high-dimensional quantum system. *Phys. Rev. Lett.*, 114:240401, 2015.

[32] J. Helsen, X. Xue, L. M. K. Vendersypen, and S. Wehner. A new class of efficient randomized benchmarking protocols, 2018. arXiv:1806.02048.

[33] J. Helsen, J. Wallman, S. T. Flammia, and S. Wehner. Multiqubit randomized benchmarking using few samples, 2017. arXiv:1701.04299.

[34] J.-P. Serre. *Linear representations of finite groups*. Springer, 1996.

[35] R. Berndt. *Representations of linear groups*. Vieweg, first edition edition, 2007.

[36] F. Casas, A. Murua, and M. Nadinic. Efficient computation of the Zassenhaus formula. *Comp. Phys. Comm.*, 183:2386–2391, 2012.

[37] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Phys. Rev. A*, 303:249–252, 2002.

[38] D. P. O'Leary and B. W. Rust. Variable projection for nonlinear least squares problems. *Computational Optimization and Applications*, 54:579–593, 2013.

[39] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Ac. Sp. Sig. Proc.*, 38:814–824, May 1990.

[40] T. K. Sarkar and O. Pereira. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *IEEE Ant. Prop. Mag.*, 37:48–55, Feb 1995.

[41] D. Potts and M. Tasche. Parameter estimation for nonincreasing exponential sums by prony-like methods. *Linear Algebra and its Applications*, 439:1024 – 1039, 2013. 17th Conference of the International Linear Algebra Society, Braunschweig, Germany, August 2011.

[42] J.-M. Papy, L. De Lathauwer, and S. Van Huffel. Common pole estimation in multi-channel exponential data modeling. *Signal Processing*, 86:846 – 858, 2006.

[43] D. S. Moore, G. P. McCabe, and B. A. Craig. *Introduction to the practice of statistics*. W. H. Freeman, 2016.

[44] J.-P. Serre. *Linear representations of finite groups*. Springer, 1996.

[45] R. Berndt. *Representations of linear groups*. Vieweg, first edition edition, 2007.

[46] F. Casas, A. Murua, and M. Nadinic. Efficient computation of the Zassenhaus formula. *Comp. Phys. Comm.*, 183:2386–2391, 2012.

[47] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Phys. Rev. A*, 303:249–252, 2002.

[48] 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*, 4:011050, 2014.

[49] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, 2012.

[50] J. J. Wallman and S. T. Flammia. Randomized benchmarking with confidence. *New. J. Phys.*, 16:103032, 2014.

[51] J. Helsen, J. Wallman, S. T. Flammia, and S. Wehner. Multi-qubit randomized benchmarking using few samples, 2017. arXiv:1701.04299.

[52] D. P. O'Leary and B. W. Rust. Variable projection for nonlinear least squares problems. *Computational Optimization and Applications*, 54:579–593, 2013.

[53] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Ac. Sp. Sig. Proc.*, 38:814–824, May 1990.

[54] T. K. Sarkar and O. Pereira. Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *IEEE Ant. Prop. Mag.*, 37:48–55, Feb 1995.

[55] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38:814–824, May 1990.

[56] D. Potts and M. Tasche. Parameter estimation for nonincreasing exponential sums by prony-like methods. *Linear Algebra and its Applications*, 439:1024 – 1039, 2013. 17th Conference of the International Linear Algebra Society, Braunschweig, Germany, August 2011.

[57] J.-M. Papy, L. De Lathauwer, and S. Van Huffel. Common pole estimation in multi-channel exponential data modeling. *Signal Processing*, 86:846 – 858, 2006.

[58] J. Helsen, X. Xue, L. M. K. Vendersypen, and S. Wehner. A new class of efficient randomized benchmarking protocols, 2018. arXiv:1806.02048.

[59] D. S. Moore, G. P. McCabe, and B. A. Craig. *Introduction to the practice of statistics*. W. H. Freeman, 2016.