Delft University of Technology

A new class of efficient randomized benchmarking protocols

Helsen, Jonas; Xue, Xiao; Vandersypen, Lieven M.K.; Wehner, Stephanie

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ARTICLE   OPEN

A new class of efficient randomized benchmarking protocols

Jonas Helsen [ID][1], Xiao Xue[1,2], Lieven M. K. Vandersypen [ID][1,2] and Stephanie Wehner[1]

Randomized benchmarking is a technique for estimating the average fidelity of a set of quantum gates. However, if this gateset is not the multi-qubit Clifford group, robustly extracting the average fidelity is difficult. Here, we propose a new method based on representation theory that has little experimental overhead and robustly extracts the average fidelity for a broad class of gatesets. We apply our method to a multi-qubit gateset that includes the *T*-gate, and propose a new interleaved benchmarking protocol that extracts the average fidelity of a two-qubit Clifford gate using only single-qubit Clifford gates as reference.

INTRODUCTION

Randomized benchmarking[1–7] is arguably the most prominent experimental technique for assessing the quality of quantum operations in experimental quantum computing devices.[4,8–13] Key to the wide adoption of randomized benchmarking are its scalability with respect to the number of qubits and its insensitivity to errors in state preparation and measurement. It has also recently been shown to be insensitive to variations in the error associated to different implemented gates.[14–16]

The randomized benchmarking protocol is defined with respect to a gateset G, a discrete collection of quantum gates. Usually, this gateset is a group, such as the Clifford group.[2] The goal of randomized benchmarking is to estimate the average fidelity[17] of this gateset.

Randomized benchmarking is performed by randomly sampling a sequence of gates of a fixed length m from the gateset G. This sequence is applied to an initial state ρ, followed by a global inversion gate such that in the absence of noise the system is returned to the starting state. Then the overlap between the output state and the initial state is estimated by measuring a two-component POVM $\{Q, 1 - Q\}$. This is repeated for many sequences of the same length m and the outputs are averaged, yielding a single average survival probability p_m. Repeating this procedure for various sequence lengths m yields a list of probabilities $\{p_m\}_m$.

Usually G is chosen to be the Clifford group. It can then be shown (under the assumption of gate-independent CPTP noise)[2] that the data $\{p_m\}_m$ can be fitted to a single exponential decay of the form

$$p_m \approx_{\text{fit}} A + Bf^m \tag{1}$$

where A, B depend on state preparation and measurement, and the quality parameter f only depends on how well the gates in the gateset G are implemented. This parameter f can then be straightforwardly related to the average fidelity F_{avg}.[2] The fitting relation Eq. (1) holds intuitively because averaging over all elements of the Clifford group effectively depolarizes the noise affecting the input state ρ. This effective depolarizing noise then accretes exponentially with sequence length m.

However it is possible, and desirable, to perform randomized benchmarking on gatesets that are not the Clifford group, and a wide array of proposals for randomized benchmarking using non-Clifford gatesets appear in the literature.[18–24] The most prominent use case is benchmarking a gateset G that includes the vital *T*-gate[18,19,22] which, together with the Clifford group, forms a universal set of gates for quantum computing.[17] Another use case is simultaneous randomized benchmarking,[23] which extracts information about crosstalk and unwanted coupling between neighboring qubits by performing randomized benchmarking on the gateset consisting of single qubit Clifford gates on all qubits. In these cases, and in other examples of randomized benchmarking with non-Clifford gatesets,[20,22,23] the fitting relation Eq. (1) does not hold and must instead be generalized to

$$p_m \approx_{\text{fit}} \sum_{\lambda \in R_G} A_\lambda f_\lambda^m, \tag{2}$$

where R_G is an index set that only depends on the chosen gateset, the f_λ are general 'quality parameters' that only depend on the gates being implemented and the A_λ prefactors depend only on SPAM (when the noise affecting the gates is trace preserving there will be a $\lambda \in R_G$ -corresponding to the trivial subrepresentation- such that $f_\lambda = 1$, yielding the constant offset seen in Eq. (1)). The above holds because averaging over sequences of elements of these non-Clifford groups averaging does not fully depolarize the noise. Rather the system state space will split into several 'sectors' labeled by λ, with a different depolarization rate, set by f_λ, affecting each sector. The interpretation of the parameters f_λ varies depending on the gateset G. In the case of simultaneous randomized benchmarking[23] they can be interpreted as a measure of crosstalk and unwanted coupling between neighboring qubits. For other gatesets an interpretation is not always available. However, as was pointed out for specific gatesets in ref. [18–20,22] and for general finite groups in ref.,[21] the parameters f_λ can always be jointly related (see Eq. (5)) to the average fidelity F_{avg} of the gateset G. This means that in theory randomized benchmarking can extract the average fidelity of a gateset even when it is not the Clifford group.

However in practice the multi-parameter fitting problem given by Eq. (2) is difficult to perform, with poor confidence intervals

around the parameters f_λ unless impractically large amounts of data are gathered. More fundamentally it is, even in the limit of infinite data, impossible to associate the estimates from the fitting procedure to the correct decay channel in Eq. (2) and thus to the correct f_λ, making it impossible to reliably reconstruct the average fidelity of the gateset.

In the current literature on non-Clifford randomized benchmarking, with the notable exception of ref.,[22] this issue is sidestepped by performing randomized benchmarking several times using different input states ρ_λ that are carefully tuned to maximize one of the prefactors A_λ while minimizing the others. This is unsatisfactory for several reasons: (1) the accuracy of the fit now depends on the preparation of ρ_λ, undoing one of the main advantages of randomized benchmarking over other methods such as direct fidelity estimation,[25] and (2) it is, for more general gatesets, not always clear how to find such a maximizing state ρ_λ. These problems aren't necessarily prohibitive for small numbers of qubits and/or exponential decays (see for instance[26]) but they do limit the practical applicability of current non-Clifford randomized benchmarking protocols on many qubits and more generally restrict which groups can practically be benchmarked.

Here, we propose an adaptation of the randomized benchmarking procedure, which we call character randomized benchmarking, which solves the above problems and allows reliable and efficient extraction of average fidelities for gatesets that are not the Clifford group. We begin by discussing the general method, before applying it to specific examples. Finally, we discuss using character randomized benchmarking in practice and argue the new method does not impose significant experimental overhead. Previous adaptations of randomized benchmarking, as discussed in[8,27,28] and in particular[22] (where the idea of projecting out exponential decays was first proposed for a single qubit protocol), can be regarded as special cases of our method.

RESULTS

In this section, we present the main result of this paper: the character randomized benchmarking protocol, which leverages techniques from character theory[29] to isolate the exponential decay channels in Eq. (2). One can then fit these exponential decays one at a time, obtaining the quality parameters f_λ. We emphasize that the data generated by character randomized benchmarking can always be fitted to a single exponential, even if the gateset being benchmarked is not the Clifford group. Moreover, our method retains its validity in the presence of leakage, which also causes deviations from single exponential behavior for standard randomized benchmarking[14] (even when the gateset is the Clifford group).

For the rest of the paper, we will use the Pauli Transfer Matrix (PTM) representation of quantum channels (This representation is also sometimes called the Liouville representation or affine representation of quantum channels[30,31]). Key to this representation is the realization that the set of normalized non-identity Pauli matrices $\sigma_{\mathbf{q}}$ on q qubits, together with the normalized identity $\sigma_0 := 2^{-q/2}\mathbb{1}$ forms an orthonormal basis (with respect to the trace inner product) of the Hilbert space of Hermitian matrices of dimension 2^q. Density matrices ρ and POVM elements Q can then be seen as vectors and co-vectors expressed in the basis $\{\sigma_0\} \cup \sigma_{\mathbf{q}}$, denoted $|\rho\rangle\rangle$ and $\langle\langle Q|$ respectively. Quantum channels $\mathcal{E}$[32] are then matrices (we will denote a channel and its PTM representation by the same letter) and we have $\mathcal{E}|\rho\rangle\rangle = |\mathcal{E}(\rho)\rangle\rangle$. Composition of channels $\mathcal{E}, \mathcal{F}$ corresponds to multiplication of their PTM representations, that is $|\mathcal{E} \circ \mathcal{F}(\rho)\rangle\rangle = \mathcal{E}\mathcal{F}|\rho\rangle\rangle$. Moreover, we can write expectation values as bra-ket inner products, i.e. $\langle\langle Q|\mathcal{E}|\rho\rangle\rangle = \mathrm{Tr}(Q\mathcal{E}(\rho))$. The action of a unitary G on a matrix ρ is denoted $\mathcal{G}$, i.e. $\mathcal{G}|\rho\rangle\rangle = |G\rho G^\dagger\rangle\rangle$ and we denote its noisy implementation by $\tilde{\mathcal{G}}$. For a more expansive review of the PTM representation, see Section I.2 in the Supplementary Methods.

We will, for ease of presentation, also assume *gate-independent noise*. This means we assume the existence of a CPTP map $\mathcal{E}$ such that $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ for all $G \in G$. We however emphasize that our protocol remains functional even in the presence of gate-dependent noise. We provide a formal proof of this, generalizing the modern treatment of standard randomized benchmarking with gate-dependent noise,[14] in the Methods section.

Standard randomized benchmarking

Let's first briefly recall the ideas behind standard randomized benchmarking. Subject to the assumption of gate-independent noise, the average survival probability p_m of the standard randomized benchmarking procedure over a gateset G (with input state ρ and measurement POVM $\{Q, 1 - Q\}$) with sequence length m can be written as: ref.[2]

$$p_m = \langle\langle Q| \left(\mathop{\mathbb{E}}_{G \in G} \mathcal{G}^\dagger \mathcal{E}\mathcal{G} \right)^m |\rho\rangle\rangle. \tag{3}$$

where $\mathbb{E}_{G \in G}$ denotes the uniform average over G. The key insight to randomized benchmarking is that $\mathcal{G}$ is a *representation* (for a review of representation theory see section I.1 in the Supplementary Methods) of $G \in G$. This representation will not be irreducible but will rather decompose into irreducible subrepresentations, that is $\mathcal{G} = \oplus_{\lambda \in R_G}\phi_\lambda(G)$ where R_G is an index set and ϕ_λ are irreducible representations of G which we will assume to all be mutually inequivalent. Using Schur's lemma, a fundamental result in representation theory, we can write Eq. (3) as

$$p_m = \sum_\lambda \langle\langle Q|\mathcal{P}_\lambda|\rho\rangle\rangle f_\lambda^m \tag{4}$$

where $\mathcal{P}_\lambda$ is the orthogonal projector onto the support of ϕ_λ (note that this is a superoperator) and $f_\lambda := \mathrm{Tr}(\mathcal{P}_\lambda\mathcal{E})/\mathrm{Tr}(\mathcal{P}_\lambda)$ is the quality parameter associated to the representation ϕ_λ (note that the trace is taken over superoperators). This reproduces Eq. (2). A formal proof of Eq. (4) can be found in the Supplementary Methods and in ref.[21] The average fidelity of the gateset G can then be related to the parameters f_λ as

$$F_{\mathrm{avg}} = \frac{2^{-q} \sum_{\lambda \in R_G} \mathrm{Tr}(\mathcal{P}_\lambda)f_\lambda}{2^q + 1}. \tag{5}$$

Note again that R_G includes the trivial subrepresentation carried by $|1\rangle\rangle$, so when $\mathcal{E}$ is a CPTP map there is a $\lambda \in R_G$ for which $f_\lambda = 1$. See Lemma's 4 and 5 in the Supplementary Methods for a proof of Eq. (5)

Character randomized benchmarking

Now we present our new method called character randomized benchmarking. For this we make use of concepts from the character theory of representations.[29] Associated to any representation $\hat{\phi}$ of a group $\hat{G}$ is a character function $\chi_{\hat{\phi}} : \hat{G} \to \mathbb{R}$, from the group to the real numbers (Generally the character function is a map to the complex numbers, but in our case it is enough to only consider real representations). Associated to this character function is the following projection formula:[29]

$$\mathop{\mathbb{E}}_{\hat{G} \in \hat{G}} \chi_{\hat{\phi}}(\hat{G})\hat{\mathcal{G}} = \frac{1}{|\hat{\phi}|}\mathcal{P}_{\hat{\phi}}, \tag{6}$$

where $\mathcal{P}_{\hat{\phi}}$ is the projector onto the support of all subrepresentations of $\hat{\mathcal{G}}$ equivalent to $\hat{\phi}$ and $|\hat{\phi}|$ is the dimension of the representation $\hat{\phi}$. We will leverage this formula to adapt the randomized benchmarking procedure in a way that singles out a particular exponential decay f_λ^m in Eq. (2).

We begin by choosing a group G. We will call this group the 'benchmarking group' going forward and it is for this group/gateset that we will estimate the average fidelity. In general we will have that $\mathcal{G} = \oplus_{\lambda \in R_G}\phi_\lambda(G)$ where R_G is an index set and ϕ_λ are irreducible representations of G which we will assume to all be

1. Choose a state ρ and a two-component POVM $\{Q, \mathbb{1} - Q\}$ such that $\mathrm{Tr}(Q\mathcal{P}_{\hat{\phi}}(\rho))$ is large.

2. Sample $\vec{G} = G_1, \ldots, G_m$ uniformly at random from G

3. Sample $\hat{G}$ uniformly at random from $\hat{\mathsf{G}}$

4. Prepare the state ρ and apply the gates $(G_1\hat{G}), G_2, \ldots G_m$

5. Compute the inverse $G_{\mathrm{inv}} = (G_m \cdots G_1)^{\dagger}$ and apply it (note that $\hat{G}$ is not inverted)

6. Estimate the weighted 'survival probability'
$$k_m^{\lambda'}(\vec{G}, \hat{G}) = |\hat{\phi}|\chi_{\hat{\phi}}(\hat{G})\langle\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\mathcal{G}}_m \cdots \widetilde{(\mathcal{G}_1\hat{\mathcal{G}})}|\rho\rangle\rangle$$

7. Repeat for sufficient $\hat{G} \in \hat{\mathsf{G}}$ to estimate the average $k_m^{\lambda'}(\vec{G}) = \mathbb{E}_{\hat{G}}(k_m^{\lambda'}(\vec{G}, \hat{G}))$

8. Repeat for sufficient $\vec{G}$ to estimate the average $k_m^{\lambda'} = \mathbb{E}_{\vec{G}}(k_m^{\lambda'}(\vec{G}))$

9. Repeat for sufficient different m to fit to the exponential function $Af_{\lambda'}^m$ to obtain $f_{\lambda'}$

Fig. 1 The character randomized benchmarking protocol. Note the inclusion of the gate $\hat{G}$ and the average over the character function $\chi_{\hat{\phi}}$, which form the key ideas behind character randomized benchmarking. Note also that this extra gate $\hat{G}$ is compiled into the sequence of gates $(G_1, \ldots, G_m)$ and thus does not result in extra noise

mutually inequivalent (It is straightforward to extend character randomized benchmarking to also cover the presence of equivalent irreducible subrepresentation. However do not make this extension explicit here in the interest of simplicity). Now fix a $\lambda' \in R_\mathsf{G}$. $f_{\lambda'}$ is the quality parameter associated to a specific subrepresentation $\phi_{\lambda'}$ of $\mathcal{G}$. Next consider a group $\hat{\mathsf{G}} \subset \mathsf{G}$ such that the PTM representation $\hat{\mathcal{G}}$ has a subrepresentation $\hat{\phi}$, with character function $\chi_{\hat{\phi}}$, that has support inside the representation $\phi_{\lambda'}$ of G, i.e. $\mathcal{P}_{\hat{\phi}} \subset \mathcal{P}_{\lambda'}$ where $\mathcal{P}_{\lambda'}$ is again the projector onto the support of $\phi_{\lambda'}$. We will call this group $\hat{\mathsf{G}}$ the character group. Note that such a pair $\hat{\mathsf{G}}, \hat{\phi}$ always exists; we can always choose $\hat{\mathsf{G}} = \mathsf{G}$ and $\hat{\phi} = \phi_{\lambda'}$. However other natural choices often exist, as we shall see when discussing examples of character randomized benchmarking. The idea behind the character randomized benchmarking protocol, described in Fig. 1, is now to effectively construct Eq. (6) by introducing the application of an extra gate $\hat{G}$ drawn at random from the character group $\hat{\mathsf{G}}$ into the standard randomized benchmarking protocol. In practice this gate will not be actively applied but must be compiled into the gate sequence following it, thus not resulting in extra noise (this holds even in the case of gate-dependent noise, see Methods).

This extra gate $\hat{G} \in \hat{\mathsf{G}}$ is not included when computing the global inverse $G_{\mathrm{inv}} = (G_1 \ldots G_m)^{\dagger}$. The average over the elements of $\hat{\mathsf{G}}$ is also weighted by the character function $\chi_{\hat{\phi}}$ associated to the representation $\hat{\phi}$ of $\hat{\mathsf{G}}$. Similar to eq. (3) we can rewrite the uniform average over all $\vec{G} \in \mathsf{G}^{\times m}$ and $\hat{G} \in \hat{\mathsf{G}}$ as

$$k_m^{\lambda'} = |\hat{\phi}|\langle\langle Q| \left[\mathbb{E}_{G \in \mathsf{G}} \mathcal{G}^{\dagger}\mathcal{E}\mathcal{G} \right]^m \mathbb{E}_{\hat{G} \in \hat{\mathsf{G}}} \chi_{\hat{\phi}}(\hat{G})\hat{\mathcal{G}}|\rho\rangle\rangle.$$

Using the character projection formula (Eq. (6)), the linearity of quantum mechanics, and the standard randomized benchmarking representation theory formula (Eq. (4)) we can write this as

$$k_m^{\lambda'} = \sum_{\lambda \in R_\mathsf{G}} \langle\langle Q|\mathcal{P}_\lambda\mathcal{P}_{\hat{\phi}}|\rho\rangle\rangle f_\lambda^m = \langle\langle Q|\mathcal{P}_{\hat{\phi}}|\rho\rangle\rangle f_{\lambda'}^m \quad (7)$$

since we have chosen $\hat{\mathsf{G}}$ and $\hat{\phi}$ such that $\mathcal{P}_{\hat{\phi}} \subset \mathcal{P}_{\lambda'}$. This means the character randomized benchmarking protocol isolates the exponential decay associated to the quality parameter $f_{\lambda'}$ independent of state preparation and measurement. We can now extract $f_{\lambda'}$ by fitting the data-points $k_m^{\lambda'}$ to a single exponential of the form $Af_{\lambda'}^m$. Note that this remains true even if $\mathcal{E}$ is not trace-preserving, i.e. the implemented gates experience leakage. Repeating this procedure

for all $\lambda' \in R_\mathsf{G}$ (choosing representations $\hat{\phi}$ of $\hat{\mathsf{G}}$ such that $\mathcal{P}_{\hat{\phi}} \subset \mathcal{P}_{\lambda'}$) we can reliably estimate all quality parameters f_λ associated to randomized benchmarking over the group G. Once we have estimated all these parameters we can use Eq. (5) to obtain the average fidelity of the gateset G.

DISCUSSION

We will now discuss several examples of randomized benchmarking experiments where the character randomized benchmarking approach is beneficial. The first example, benchmarking T-gates, is taken from the literature[18] while the second one, performing interleaved benchmarking on a 2-qubit gate using only single qubit gates a reference, is a new protocol. We have also implemented this last protocol to characterize a CPHASE gate between spin qubits in Si/SiGe quantum dots, see ref. [33]

Benchmarking T-gates

The most common universal gateset considered in the literature is the Clifford + T gateset.[17] The average fidelity of the Clifford gates can be extracted using standard randomized benchmarking over the Clifford group, but to extract the average fidelity of the T gate a different approach is needed. Moreover one would like to characterize this gate in the context of larger circuits, meaning that we must find a family of multi-qubit groups that contains the T gate. One choice is to perform randomized benchmarking over the group T_q generated by the CNOT gate between all pairs of qubits (in both directions), Pauli X on all qubits and T gates on all qubits (another choice would be to use dihedral randomized benchmarking[22] but this is limited to single qubit systems, or to use the interleaved approach proposed in ref. [24]). This group is an example of a CNOT-dihedral group and its use for randomized benchmarking was investigated in.[18] There it was derived that the PTM representation of the group T_q decomposes into 3 irreducible subrepresentations ϕ_1, ϕ_2, ϕ_3 with associated quality parameters f_1, f_2, f_3 and projectors

$$\mathcal{P}_1 = |\sigma_0\rangle\rangle\langle\langle\sigma_0|, \quad \mathcal{P}_2 = \sum_{\sigma \in \mathcal{Z}_q} |\sigma\rangle\rangle\langle\langle\sigma|, \quad \mathcal{P}_3 = \sum_{\sigma \in \sigma_q/\mathcal{Z}_q} |\sigma\rangle\rangle\langle\langle\sigma|,$$

where σ_0 is the normalized identity, σ_q is the set of normalized Pauli matrices and $\mathcal{Z}_q$ is the subset of the normalized Pauli matrices composed only of tensor products of Z and $\mathbb{1}$. Noting that $f_1 = 1$ if the implemented gates $\widetilde{\mathcal{G}}$ are CPTP we must estimate f_2 and f_3 in order to estimate the average fidelity of T_q. Using standard randomized benchmarking this would thus lead to a two-decay, four-parameter fitting problem, but using character randomized benchmarking we can fit f_2 and f_3 separately. Let's say we want to estimate f_2, associated to ϕ_2, using character randomized benchmarking. In order to perform character randomized benchmarking we must first choose a character group $\hat{\mathsf{G}}$. A good choice for $\hat{\mathsf{G}}$ is in this case the Pauli group P_q. Note that $\mathsf{P}_q \subset T_q$ since $T^4 = Z$ the Pauli Z matrix.

Having chosen $\hat{\mathsf{G}} = \mathsf{P}_q$ we must also choose an irreducible subrepresentation $\hat{\phi}$ of the PTM representation of the Pauli group P_q such that $\mathcal{P}_{\hat{\phi}}\mathcal{P}_2 = \mathcal{P}_{\hat{\phi}}$. As explained in detail in section V.I in the Supplementary Methods the PTM representation of the Pauli group has 2^q irreducible inequivalent subrepresentations of dimension one. These representations ϕ_σ are each associated to an element $\sigma \in \{\sigma_0\} \cup \boldsymbol{\sigma_q}$ of the Pauli basis. Concretely we have that the projector onto the support of ϕ_σ is given by $\mathcal{P}_\sigma = |\sigma\rangle\rangle\langle\langle\sigma|$. This means that, to satisfy $\mathcal{P}_{\hat{\phi}}\mathcal{P}_2 = \mathcal{P}_{\hat{\phi}}$ we have to choose $\hat{\phi} = \phi_\sigma$ with $\sigma \in \mathcal{Z}_q$. One could for example choose σ proportional to $Z^{\otimes q}$. The character associated to the representation ϕ_σ is $\chi_{\phi_\sigma}(P) = (-1)^{\langle P, \sigma\rangle}$ where $\langle P, \sigma\rangle = 1$ if and only if P and σ anti-commute and zero otherwise (we provided a proof of this fact in section V.1 of the Supplementary Methods). Hence the character randomized benchmarking experiment with

benchmarking group T_q, character group P_q and subrepresentation $\hat{\phi} = \phi_\sigma$ produces data that can be described by

$$k_m^2 = \langle\langle Q|\sigma\rangle\rangle\langle\langle\sigma|\rho\rangle\rangle f_2^m, \tag{8}$$

allowing us to reliably extract the parameter f_2. We can perform a similar experiment to extract f_3, but we must instead choose $\sigma \in \boldsymbol{\sigma}_q \setminus \mathcal{Z}$. A good choice would for instance be σ proportional to $X^{\otimes q}$.

Having extracted f_2 and f_3 we can then use Eq. (5) to obtain the average fidelity of the gateset T_q as:[18]

$$F_{\text{avg}} = \frac{2^q - 1}{2^q}\left(1 - \frac{f_2 + 2^q f_3}{2^q + 1}\right) \tag{9}$$

Finally we would like to note that in order to get good signal one must choose ρ and Q appropriately. The correct choice is suggested by Eq. (7). For instance, if when estimating f_2 as above we choose σ proportional to $Z^{\otimes q}$ we must then choose $Q = \frac{1}{2}(1 + Z^{\otimes 2})$ and $\rho = \frac{1}{d}(1 + Z^{\otimes 2})$. This corresponds to the even parity eigenspace (in the computational basis).

2-for-1 interleaved benchmarking

The next example is a new protocol, which we call 2-for-1 interleaved randomized benchmarking. It is a way to perform interleaved randomized benchmarking[34] of a 2-qubit Clifford gate C using only single qubit Clifford gates as reference gates. The advantages of this are (1) lower experimental requirements and (2) a higher reference gate fidelity relative to the interleaved gate fidelity allows for a tighter estimate of the average fidelity of the interleaved gate (assuming single qubit gates have higher fidelity than two qubit gates). This latter point is related to an oft overlooked drawback of interleaved randomized benchmarking, namely that it does not yield a direct estimate of the average fidelity $F(C)$ of the interleaved gate C but only gives upper and lower bounds on this fidelity. These upper and lower bounds moreover depend[34,35] on the fidelity of the reference gates and can be quite loose if the fidelity of the reference gates is low. To illustrate the advantages of this protocol we have performed a simulation comparing it to standard interleaved randomized benchmarking (details can be found in section V.2 in the Supplementary Methods). Following recent single qubit randomized benchmarking and Bell state tomography results in spin qubits in Si/SiGe quantum dots[36–38] we assumed single qubit gates to have a fidelity of $F_{\text{avg}}^{(1)} = 0.987$ and two-qubit gates to have a fidelity of $F_{\text{avg}}(C) = 0.898$. Using standard interleaved randomized benchmarking[34] we can guarantee (using the optimal bounds of ref. [35]) that the fidelity of the interleaved gate is lower bounded by $F_{\text{avg}}^{\text{int}} \approx 0.62$ while using 2-for-1 interleaved randomized benchmarking we can guarantee that the fidelity of interleaved gate is lower bounded by $F_{\text{avg}}(C) \approx 0.79$, a significant improvement that is moreover obtained by a protocol requiring less experimental resources. On top of this the 2-for-1 randomized benchmarking protocol provides strictly more information than simply the average fidelity, we can also extract a measure of correlation between the two qubits, as per.[23] In another paper[33] we have used this protocol to characterize a CPHASE gate between spin qubits in Si/SiGe quantum dots.

An interleaved benchmarking experiment consists of two stages, a reference experiment and an interleaved experiment. The reference experiment for 2-for-1 interleaved randomized benchmarking consists of character randomized benchmarking using 2 copies of the single-qubit Clifford group $G = C_1^{\otimes 2}$ as the benchmarking group (this is also the group considered in simultaneous randomized benchmarking[23]). The PTM representation of $C_1^{\otimes 2}$ decomposes into four irreducible subrepresentations and thus the fitting problem of a randomized benchmarking experiment over this group involves 4 quality parameters f_w indexed by $w = (w_1, w_2) \in \{0, 1\}^{\times 2}$. The projectors onto the

associated irreducible representations ϕ_w are

$$\mathcal{P}_w = \sum_{\sigma \in \sigma_w} |\sigma\rangle\rangle\langle\langle\sigma| \tag{10}$$

where σ_w is the set of normalized 2-qubit Pauli matrices that have non-identity Pauli matrices at the i'th tensor factor if and only if $w_i = 1$. To perform character randomized benchmarking we choose as character group $\hat{G} = P_2$ the 2-qubit Pauli group. For each $w \in \{0, 1\}^{\times 2}$ we can isolate the parameter f_w by correctly choosing a subrepresentation ϕ_σ of the PTM representation of P_2. Recalling that $\mathcal{P}_\sigma = |\sigma\rangle\rangle\langle\langle\sigma|$ we can choose $\hat{\phi} = \phi_\sigma$ for $\sigma = (Z_1^w \otimes Z_2^w)/2$ to isolate the parameter f_w for $w = (w_1, w_2) \in \{0, 1\}^{\times 2}$. We give the character functions associated to these representation in section V.2 of the Supplementary Methods. Once we have obtained all quality parameters f_w we can compute the average reference fidelity F_{ref} using Eq. (5).

The interleaved experiment similarly consists of a character randomized benchmarking experiment using $G = C_1^{\otimes 2}$ but for every sequence $\vec{G} = (G_1, \dots, G_m)$ we apply the sequence $(G_1, C, G_2, \dots, C, G_m)$ instead, where C is a 2-qubit interleaving gate (from the 2-qubit Clifford group). Note that we must then also invert this sequence (with C) to the identity.[34] Similarly choosing $\hat{G} = P_2$ we can again isolate the parameters f_w and from these compute the 'interleaved fidelity' F_{int}. Using the method detailed in ref. [35] we can then calculate upper and lower bounds on the average fidelity $F_{\text{avg}}(C)$ of the gate C from the reference fidelity F_{ref} and the interleaved fidelity F_{int}. Note that it is not trivial that the interleaved experiment yields data that can be described by a single exponential decay, we will discuss this in greater detail in the methods section.

Finally we would like to note that the character benchmarking protocol can be used in many more scenarios than the ones outlined here. Character randomized benchmarking is versatile enough that when we want to perform randomized benchmarking we can consider first what group is formed by the native gates in our device and then use character benchmarking to extract gate fidelities from this group directly, as opposed to carefully compiling the Clifford group out of the native gates which would be required for standard randomized benchmarking. This advantage is especially pronounced when the native two-qubit gates are not part of the Clifford group, which is the case for e.g. the $\sqrt{\text{SWAP}}$ gate.[39,40]

METHODS

In this section will discuss three things: (1) The statistical behavior and scalability of character randomized benchmarking, (2) the robustness of character randomized benchmarking against gate-dependent noise, and (3) the behavior of interleaved character randomized benchmarking, and in particular 2-for-1 interleaved benchmarking.

First we will consider whether the character randomized benchmarking protocol is efficiently scalable with respect to the number of qubits (like standard randomized benchmarking) and whether the character randomized benchmarking protocol remains practical when only a finite amount of data can be gathered (this last point is a sizable line of research for standard randomized benchmarking[6,28,30,41]).

Scalability of character randomized benchmarking

The resource cost (the number of experimental runs that must be performed to obtain an estimate of the average fidelity) of character randomized benchmarking can be split into two contributions: (1) The number of quality parameters f_λ associated that must be estimated (this is essentially set by $|R_G|$, the number of irreducible subrepresentations of the PTM representation of the benchmarking group G), and (2) the cost of estimating a single average k_m^λ for a fixed $\lambda' \in R_G$ and sequence length m.

The first contribution implies that for scalable character randomized benchmarking with (a uniform family of) groups G_q (w.r.t. the number of qubits q) the number of quality parameters (set by $|R_G|$) must grow polynomially with q. This means that not all families of benchmarking

groups are can be characterized by character randomized benchmarking in a scalable manner.

The second contribution, as can be seen in Fig. 1, further splits up into three components: (2a) the magnitude of $|\hat{\phi}|$, (2b) the number of random sequences $\vec{G}$ needed to estimate k_m^λ (given access to $k_m^\lambda(\vec{G})$) and (2c) the number of samples needed to estimate $k_m^\lambda(\vec{G})$ for a fixed sequence. We will now argue that the resource cost of all three components are essentially set by the magnitude of $|\hat{\phi}|$. Thus if $|\hat{\phi}|$ grows polynomially with the number of qubits then the entire resource cost does so as well. Hence a sufficient condition for scalable character randomized benchmarking is that one chooses a family of benchmarking groups where $|R_G|$ grows polynomially in q and character groups such that for the relevant subrepresentations $|\hat{\phi}|$ the dimension grows polynomially in q.

We begin by arguing (2c): The character-weighted average over the group $\hat{G}$ for a single sequence $\vec{G}$: $k_m^\lambda(\vec{G})$, involves an average over $|\hat{G}|$ elements (which will generally scale exponentially in q), but can be efficiently estimated by not estimating each character-weighted expectation value $k_m^\lambda(\vec{G}, \hat{G})$ individually but rather estimate $k_m^\lambda(\vec{G})$ directly by the following procedure

1. Sample $\hat{G} \in \hat{G}$ uniformly at random
2. Prepare the state $\mathcal{G}_{inv}\mathcal{G}_m \cdots \mathcal{G}_1\hat{G}|\rho\rangle\rangle$ and measure it once obtaining a result $b(\hat{G}) \in \{0, 1\}$
3. Compute $x(\hat{G}) := \chi_{\hat{\phi}}(\hat{G})|\hat{\phi}|b(\hat{G}) \in \{0, \chi_{\hat{\phi}}(\hat{G})|\hat{\phi}|\}$
4. Repeat sufficiently many times and compute the empirical average of $x(\hat{G})$

Through the above procedure we are directly sampling from a bounded probability distribution with mean $k_m^\lambda(\vec{G})$ that takes values in the interval $[-\chi_{\hat{\phi}}^*, \chi_{\hat{\phi}}^*]$ where $\chi_{\hat{\phi}}^*$ is the largest absolute value of the character function $\chi_{\hat{\phi}}$. Since the maximal absolute value of the character function is bounded by the dimension of the associated representation,[29] this procedure will be efficient as long as $|\hat{\phi}|$ is not too big.

For the examples given in the discussion section (with the character group being the Pauli group) the maximal character value is 1. Using standard statistical techniques[42] we can give e.g. a 99% confidence interval of size 0.02 around $k_m^\lambda(\vec{G})$ by repeating the above procedure 1769 times, which is within an order of magnitude of current experimental practice for confidence intervals around regular expectation values and moreover independent of the number of qubits q. See section VI in the Supplementary Methods for more details on this.

We now consider (2b): From the considerations above we know that $k_m^\lambda(\vec{G})$ is the mean of a set of random variables and thus itself a random variable, taking possible values in the interval $[-\chi_{\hat{\phi}}^*, \chi_{\hat{\phi}}^*]$. Hence by the same reasoning as above we see that k_m^λ, as the mean of a distribution (induced by the uniform distribution of sequences $\vec{G}$) confided to the interval $[-\chi_{\hat{\phi}}^*, \chi_{\hat{\phi}}^*]$ can be estimated using an amount of resources polynomially bounded in $|\hat{\phi}|$. We would like to note however that this estimate is probably overly pessimistic in light of recent results for standard randomized benchmarking on the Clifford group[28,30] where it was shown that the average k_m^λ over sequences $\vec{G} \in G^{\times m}$ can be estimated with high precision and high confidence using only a few hundred sequences. These results depend on the representation theoretic structure of the Clifford group but we suspect that it is possible to generalize these results at least partially to other families of benchmarking groups. Moreover any such result can be straightforwardly adapted to also hold for character randomized benchmarking. Actually making such estimates for other families groups is however an open problem, both for standard and character randomized benchmarking.

To summarize, the scalability of character randomized benchmarking depends on the properties of the families of benchmarking and character groups chosen. One should choose the benchmarking groups such that the number of exponential decays does not grow too rapidly with the number of qubits, and one should choose the character group such that the dimension of the representation being projected on does not grow too rapidly with the number of qubits.

Gate-dependent noise

Thus far we have developed the theory of character randomized benchmarking under the assumption of gate-independent noise. This is is not a very realistic assumption. Here we will generalize our framework to include gate-dependent noise. In particular we will deal with the so called 'non-Markovian' noise model. This noise model is formally specified by the existence of a function $\Phi : G \to \mathcal{S}_{2^q}$ which assigns to each element G of the group G a quantum channel $\Phi(G) = \mathcal{E}_G$. Note that this model is not the

most general, it does not take into account the possibility of time dependent effects or memory effects during the experiment. It is however much more general and realistic than the gate-independent noise model. In this section we will prove two things:

1. A character randomized benchmarking experiment always yields data that can be fitted to a single exponential decay up to a small and exponentially decreasing corrective term.
2. The decay rates yielded by a character randomized benchmarking experiment can be related to the average fidelity (to the identity) of the noise in between gates, averaged over all gates.

Both of these statements, and their proofs, are straightforward generalizations of the work of Wallman[14] which dealt with standard randomized benchmarking. We will see that his conclusion, that randomized benchmarking measures the average fidelity of noise in between quantum gates up to a small correction, generalizes to the character benchmarking case. We begin with a technical theorem, which generalizes [[14], Theorem 2] to twirls over arbitrary groups (with multiplicity-free PTM representations).

Theorem 1. *Let G be a group such that its PTM representation $\mathcal{G} = \oplus_{\lambda \in R_G} \phi_\lambda(G)$ is multiplicity-free. Denote for all λ by f_λ the largest eigenvalue of the operator $\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))$ where $\tilde{\mathcal{G}}$ is the CPTP implementation of $G \in G$. There exist Hermicity-preserving linear superoperators $\mathcal{L}, \mathcal{R}$ such that*

$$\mathbb{E}_{G \in G}(\tilde{\mathcal{G}}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{L}\mathcal{D}_G, \tag{11}$$

$$\mathbb{E}_{G \in G}(\mathcal{G}^\dagger \mathcal{R}\tilde{\mathcal{G}}) = \mathcal{D}_G \mathcal{R}, \tag{12}$$

$$\mathbb{E}_{G \in G}(\mathcal{G}\mathcal{R}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{D}_G, \tag{13}$$

where $\mathcal{D}_G$ is defined as

$$\mathcal{D}_G = \sum_\lambda f_\lambda \mathcal{P}_\lambda, \tag{14}$$

with $\mathcal{P}_\lambda$ the projector onto the representation ϕ_λ for all $\lambda \in R_G$.

Proof. Using the definition of $\mathcal{G}$ and $\mathcal{D}_G$ we can rewrite Eq. (11) as

$$\sum_\lambda \mathbb{E}_{G \in G}(\tilde{\mathcal{G}}(\mathcal{L}\mathcal{P}_\lambda)\phi_\lambda(G)^\dagger) = \sum_\lambda f_\lambda \mathcal{L}\mathcal{P}_\lambda. \tag{15}$$

This means that, without loss of generality, we can take $\mathcal{L}$ to be of the form

$$\mathcal{L} = \sum_\lambda \mathcal{L}_\lambda, \quad \mathcal{L}_\lambda \mathcal{P}_{\lambda'} = \delta_{\lambda\lambda'}\mathcal{L}_\lambda, \quad \forall \lambda'. \tag{16}$$

Similarly we can take $\mathcal{R}$ to be

$$\mathcal{R} = \sum_\lambda \mathcal{R}_\lambda, \quad \mathcal{P}_{\lambda'}\mathcal{R}_\lambda = \delta_{\lambda\lambda'}\mathcal{R}_\lambda, \quad \forall \lambda'. \tag{17}$$

This means Eqs. (11) and (12) decompose into independent pairs of equations for each λ:

$$\mathbb{E}_{G \in G}(\tilde{\mathcal{G}}\mathcal{L}_\lambda \phi_\lambda(G)^\dagger) = f_\lambda \mathcal{L}_\lambda \tag{18}$$

$$\mathbb{E}_{G \in G}(\phi_\lambda(G)^\dagger \mathcal{R}\tilde{\mathcal{G}}) = f_\lambda \mathcal{R}_\lambda. \tag{19}$$

Next we use the vectorization operator $\text{vec} : M_{2^{2q}} \to \mathbb{R}^{2^{4q}}$ mapping the PTM representations of superoperators to vectors of length $\mathbb{R}^{2^{4q}}$. This operator has the property that for all $A, B, C \in M_{2^{2q}}$ we have

$$\text{vec}(ABC) = A \otimes C^T \text{vec}(B) \tag{20}$$

where C^T is the transpose of C. Applying this to the equations Eqs. (18) and (19) and noting that $\mathcal{G}^\dagger = \mathcal{G}^T$ since $\mathcal{G}$ is a real matrix we get the eigenvalue problems equivalent to Eqs. (18) and (19),

$$\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))\text{vec}(\mathcal{L}_\lambda) = f_\lambda \text{vec}(\mathcal{L}_\lambda) \tag{21}$$

$$\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))^T \text{vec}(\mathcal{R}_\lambda) = f_\lambda \text{vec}(\mathcal{R}_\lambda). \tag{22}$$

Since we have defined f_λ to be the largest eigenvalue of $\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))$ (and equivalently of $\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))^T$) we can choose $\text{vec}(\mathcal{L})$ and $\text{vec}(\mathcal{R})$ to be the left and right eigenvectors respectively of $\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))$ associated to f_λ. Inverting the vectorization we obtain solutions to the equations Eqs. (18) and (19) and hence also Eqs. (11) and (12). To see that this solution also satisfies Eq. (13) we note first that $\mathbb{E}_{G \in G}(\mathcal{G}\mathcal{R}_\lambda \mathcal{L}_\lambda \mathcal{G}^\dagger)$ is

proportional to P_λ for any $\mathcal{R}_\lambda, \mathcal{L}_\lambda$ satisfying Eqs. (16) and (17) (by Schur's lemma). Since the eigenvectors of $\mathbb{E}_{G \in \mathsf{G}}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))$ are only defined up to a constant we can for every λ choose proportionality constants such that $\mathbb{E}_{G \in \mathsf{G}}(\mathcal{G}\mathcal{R}_\lambda \mathcal{L}_\lambda \mathcal{G}^\dagger) = f_\lambda P_\lambda$ and thus that Eq. (13) is satisfied.

Next we prove that if we perform a character randomized benchmarking experiment with benchmarking group G, character group $\hat{G}$ and subrepresentations $\hat{\phi} \subset \phi_{\lambda'}$ for some $\lambda' \in R_\mathsf{G}$, the observed data can always be fitted (up to an exponentially small correction) to a single exponential decay. The decay rate of $f_{\lambda'}$ associated to this experiment will be the largest eigenvalue of the operator $\mathbb{E}_{G \in \mathsf{G}}(\tilde{\mathcal{G}} \otimes \phi_{\lambda'}(G))$ mentioned in the theorem above. Later we will give an operational interpretation of this number. We begin by defining, for all $G \in \mathsf{G}$ a superoperator Δ_G which captures the 'gate-dependence' of the noise implementation of $\mathcal{G}$,

$$\Delta_G := \tilde{\mathcal{G}} - \mathcal{L}\mathcal{G}\mathcal{R}, \tag{23}$$

where $\mathcal{R}, \mathcal{L}$ are defined as in Theorem 1. Using this expansion we have the following theorem, which generalizes [14, Theorem 4] to character randomized benchmarking over arbitrary finite groups with multiplicity-free PTM representation.

Theorem 2. Let G be a group such that its PTM representation $\mathcal{G} = \oplus_{\lambda \in R_\mathsf{G}} \phi_\lambda(G)$ is multiplicity-free. Consider the outcome of a character randomized benchmarking experiment with benchmarking group G, character group $\hat{G}$, subrepresentations $\hat{\phi} \subset \phi_{\lambda'}$ for some $\lambda' \in R_\mathsf{G}$, and set of sequence lengths $\mathbb{M}$. That is, consider the real number

$$k_m^{\lambda'} = \mathbb{E}_{G \in \mathsf{G}} \mathbb{E}_{\hat{G} \in \hat{G}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \widetilde{\mathcal{G}_1 \mathcal{G}} | \rho \rangle\rangle \tag{24}$$

for some input state ρ and output POVM $\{Q, 1 - Q\}$ and $m \in \mathbb{M}$. This probability can be fitted to an exponential of the form

$$k_m^{\lambda'} =_{\mathrm{fit}} A f_{\lambda'}^m + \varepsilon_m, \tag{25}$$

where A is a fitting parameter, f_λ is the largest eigenvalue of the operator $\mathbb{E}_{G \in \mathsf{G}}(\tilde{\mathcal{G}} \otimes \phi_\lambda(G))$ and $\varepsilon_m \leq \delta_1 \delta_2^m$ with

$$\delta_1 = |\hat{\phi}| \max_{\hat{G} \in \hat{G}} |\chi_{\hat{\phi}}(\hat{G})| \max_{G \in \mathsf{G}} \|\Delta_G\|_\diamond, \tag{26}$$

$$\delta_2 = \mathbb{E}_{G \in \mathsf{G}} \|\Delta_G\|_\diamond, \tag{27}$$

where $\|\cdot\|_\diamond$ is the diamond norm on superoperators.[43]

Proof. We begin by expanding $\widetilde{\mathcal{G}_1 \mathcal{G}} = \mathcal{L}\mathcal{G}_1 \hat{\mathcal{G}}\mathcal{R} + \Delta_{G_1 \hat{G}}$. This gives us

$$k_m^{\lambda'} = - \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \mathcal{L}\mathcal{G}_1 \hat{\mathcal{G}}\mathcal{R} | \rho \rangle\rangle \tag{28}$$

$$+ \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \Delta_{G_1 \hat{G}} | \rho \rangle\rangle. \tag{29}$$

We now analyze the first term in Eq. (28). Using the character projection formula, the fact that $\mathcal{G}_1 = (\mathcal{G}_{\mathrm{inv}} \mathcal{G}_m \ldots \mathcal{G}_2)^\dagger$ and Eq. (11) from Theorem 1 we get

$$\mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \mathcal{L}\mathcal{G}_1 \hat{\mathcal{G}}\mathcal{R} | \rho \rangle\rangle$$

$$= \mathop{\mathbb{E}}_{G_1, \ldots, G_m \in \mathsf{G}} \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_2 \mathcal{L}\mathcal{G}_2^\dagger \ldots \mathcal{G}_{\mathrm{inv}}^\dagger \mathcal{P}_{\hat{\phi}} \mathcal{R} | \rho \rangle\rangle \tag{30}$$

$$= \mathop{\mathbb{E}}_{G_3, \ldots, G_m \in \mathsf{G}} \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_3 \mathcal{L}\mathcal{D}_\mathsf{G} \mathcal{G}_3^\dagger \ldots \mathcal{G}_{\mathrm{inv}}^\dagger \mathcal{P}_{\hat{\phi}} \mathcal{R} | \rho \rangle\rangle \tag{31}$$

$$= \langle\langle Q | \mathcal{L}\mathcal{D}_\mathsf{G}^m \mathcal{P}_{\hat{\phi}} \mathcal{R} | \rho \rangle\rangle \tag{32}$$

$$= f_{\lambda'}^m \langle\langle Q | \mathcal{L}\mathcal{P}_{\hat{\phi}} \mathcal{R} | \rho \rangle\rangle \tag{33}$$

where we used that $\mathcal{D}_\mathsf{G}$ commutes with $\mathcal{G}$ for all $G \in \mathsf{G}$ and the fact that $\mathcal{D}_\mathsf{G} \mathcal{P}_{\hat{\phi}} = f_{\lambda'} \mathcal{P}_{\hat{\phi}}$. Next we consider the second term in Eq. (28). For this we first need to prove a technical statement. We make the following

calculation for all $j \geq 2$ and $\hat{G} \in \hat{G}$:

$$\mathop{\mathbb{E}}_{G_1, \ldots, G_m \in \mathsf{G}} \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1} \mathcal{L}\mathcal{G}_j \mathcal{R} \Delta_{G_{j-1}} \ldots \Delta_{G_1 \hat{G}} \tag{34}$$

$$= \mathop{\mathbb{E}}_{G_1, \ldots, G_m \in \mathsf{G}} \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1} \mathcal{L}\mathcal{G}_{j+1}^\dagger \ldots \mathcal{G}_m^\dagger$$

$$\times \mathcal{G}_{\mathrm{inv}} \mathcal{G}_1^\dagger \ldots \mathcal{G}_{j-1}^\dagger \mathcal{R} \Delta_{G_{j-1}} \ldots \Delta_{G_1 \hat{G}} \tag{35}$$

$$= \mathop{\mathbb{E}}_{G_1, \ldots, G_m \in \mathsf{G}} \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1} \mathcal{L}\mathcal{G}_{j+1}^\dagger \ldots \mathcal{G}_m^\dagger$$

$$\times \mathcal{G}_{\mathrm{inv}} \mathcal{G}_1^\dagger \ldots \mathcal{G}_{j-1}^\dagger \mathcal{R}(\tilde{\mathcal{G}}_{j-1} - \mathcal{L}\mathcal{G}_{j-1}\mathcal{R}) \tag{36}$$

$$\times \Delta_{G_{j-2}} \ldots \Delta_{G_1 \hat{G}}$$

$$= \mathop{\mathbb{E}}_{\substack{G_1, \ldots, G_{j-1}, \\ G_{j+1}, \ldots G_m \in \mathsf{G}}} \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_{j+1} \mathcal{L}\mathcal{G}_{j+1}^\dagger \ldots \mathcal{G}_m^\dagger$$

$$\times \mathcal{G}_{\mathrm{inv}} \mathcal{G}_1^\dagger \ldots \mathcal{G}_{j-2}^\dagger (\mathcal{D}_\mathsf{G} - \mathcal{D}\mathsf{G}) \mathcal{R} \Delta_{G_{j-2}} \ldots \Delta_{G_1 \hat{G}} \tag{37}$$

$$= 0 \tag{38}$$

where we used the definition of $\Delta_{G_{j-1}}$, the fact that $G_{j-1} = (G_m \ldots G_{j+1})^\dagger G_{\mathrm{inv}} (G_1 \ldots G_{j-1})^\dagger$ and Eqs. (12) and (13). We can apply this calculation to the second term of Eq. (28) to get

$$\mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_2 \Delta_{G_1 \hat{G}} | \rho \rangle\rangle \tag{39}$$

$$= \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots (\mathcal{L}\mathcal{G}_2 \mathcal{R} + \Delta_{G_2})$$

$$\times \Delta_{G_1 \hat{G}} | \rho \rangle\rangle \tag{40}$$

$$= \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \tilde{\mathcal{G}}_{\mathrm{inv}} \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_3 \Delta_{G_2} \Delta_{G_1 \hat{G}} | \rho \rangle\rangle \tag{41}$$

$$= \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \Delta_{G_{\mathrm{inv}}} \Delta_{G_m} \ldots \Delta_{G_1 \hat{G}} | \rho \rangle\rangle \tag{42}$$

Hence we can write

$$k_m^\lambda = f_{\lambda'}^m \langle\langle Q | \mathcal{L}\mathcal{P}_{\hat{\phi}} \mathcal{R} | \rho \rangle\rangle + \varepsilon_m \tag{43}$$

with

$$\varepsilon_m = \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \Delta_{G_{\mathrm{inv}}} \Delta_{G_m} \ldots \Delta_{G_1 \hat{G}} | \rho \rangle\rangle. \tag{44}$$

We can upper bound ε_m by

$$\mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} \chi_{\hat{\phi}}(\hat{G}) |\hat{\phi}| \langle\langle Q | \Delta_{G_{\mathrm{inv}}} \Delta_{G_m} \ldots \Delta_{G_1 \hat{G}} | \rho \rangle\rangle \tag{45}$$

$$\leq \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \ldots, G_m \in \mathsf{G}}} |\chi_{\hat{\phi}}(\hat{G})| |\hat{\phi}| \|\Delta_{G_{\mathrm{inv}}}\|_\diamond \|\Delta_{G_m}\|_\diamond \cdots \|\Delta_{G_1 \hat{G}}\|_\diamond \tag{46}$$

$$\leq \max_{\hat{G} \in \hat{G}} |\chi_{\hat{\phi}}(\hat{G})| |\hat{\phi}| \max_{G \in \mathsf{G}} \|\Delta_G\|_\diamond \left(\mathbb{E}_{G \in \mathsf{G}} \|\Delta_G\|_\diamond \right)^m. \tag{47}$$

Setting

$$\delta_1 = |\hat{\phi}| \left(\max_{\hat{G} \in \hat{G}} |\chi_{\hat{\phi}}(\hat{G})| \right) \left(\max_{G \in G} \|\Delta_G\|_\diamond \right) \tag{48}$$

$$\delta_2 = \mathop{\mathbb{E}}_{G \in G} \|\Delta_G\|_\diamond \tag{49}$$

we complete the proof.

In[14] it was shown that δ_2 is small for realistic gate-dependent noise. This implies that for large enough m the outcome of a character randomized benchmarking experiment can be described by a single exponential decay (up to a small, exponentially decreasing factor). The rate of decay $f_{\lambda'}$ can be related to the largest eigenvalue of the operator $\mathbb{E}_{G \in G}(\tilde{\mathcal{G}} \otimes \phi_{\lambda'}(G))$. We can interpret this rate of decay following Wallman[14] by setting w.l.o.g. $\tilde{\mathcal{G}} = \mathcal{L}_G \mathcal{G} \mathcal{R}$ where $\mathcal{R}$ is defined as in Theorem 1 and is invertible (we can always render $\mathcal{R}$ invertible by an arbitrary small perturbation). Now consider from $\tilde{\mathcal{G}} = \mathcal{L}_G \mathcal{G} \mathcal{R}$ and the invertibility of $\mathcal{R}$:

$$\mathop{\mathbb{E}}_{G \in G} \text{Tr}(\mathcal{G}^\dagger \mathcal{R} \tilde{\mathcal{G}} \mathcal{R}^{-1}) = \mathop{\mathbb{E}}_{G \in G} \text{Tr}(\mathcal{G}^\dagger \mathcal{R} \mathcal{L}_G \mathcal{G} \mathcal{R} \mathcal{R}^{-1}) \tag{50}$$

$$= \mathop{\mathbb{E}}_{G \in G} \text{Tr}(\mathcal{R} \mathcal{L}_G) \tag{51}$$

and moreover from Eq. (12):

$$\mathop{\mathbb{E}}_{G \in G} \text{Tr}(\mathcal{G}^\dagger \mathcal{R} \tilde{\mathcal{G}} \mathcal{R}^{-1}) = \sum_{\lambda \in R_G} f_\lambda \text{Tr}(\mathcal{P}_\lambda). \tag{52}$$

From this we can consider the average fidelity of noise *between gates* (the map $(\mathcal{R} \mathcal{L}_G)$ averaged over all gates:

$$\mathop{\mathbb{E}}_{G \in G} F_{\text{avg}}(\mathcal{R} \mathcal{L}_G) = \mathop{\mathbb{E}}_{G \in G} \frac{2^{-q} \text{Tr}(\mathcal{R} \mathcal{L}_G) + 1}{2^q + 1} \tag{53}$$

$$= \frac{2^{-q} \sum_{\lambda \in R_G} f_\lambda \text{Tr}(\mathcal{P}_\lambda) + 1}{2^q + 1}. \tag{54}$$

Hence can interpret the quality parameters given by character randomized benchmarking as characterizing the average noise in between gates, extending the conclusion reached in[14] for standard randomized benchmarking to character randomized benchmarking. In ref. [16] an alternative interpretation of the decay rate of randomized benchmarking in the presence of gate dependent noise is given in terms of Fourier transforms of matrix valued group functions. One could recast the above analysis for character randomized benchmarking in this language as well but we do not pursue this further here.

Interleaved character randomized benchmarking

In the main text we proposed 2-for-1 interleaved randomized benchmarking, a form of character interleaved randomized benchmarking. More generally we can consider performing interleaved character randomized benchmarking with a benchmarking group G, a character group $\hat{G}$, and an interleaving gate C. However it is not obvious that the interleaved character randomized benchmarking procedure (for arbitrary G and C) always yields data that can be fitted to a single exponential such that the average fidelity can be extracted. Here we will justify this behavior subject to an assumption on the relation between the interleaving gate C and the benchmarking group G which we expect to be quite general. This relation is phrased in terms of what we call the 'mixing matrix' of the group G and gate C. This matrix, which we denote by M, has entries

$$M_{\lambda,\hat{\lambda}} = \frac{1}{\text{Tr}(\mathcal{P}_\lambda)} \text{Tr}(\mathcal{P}_\lambda \mathcal{C} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger) \tag{55}$$

for $\lambda, \lambda' \in R'_G = R_G \backslash \{\text{id}\}$ with ϕ_{id} the trivial subrepresentation of the PTM representation of G carried by $|1\rangle\rangle$ and where $\mathcal{P}_\lambda$ is the projector onto the subrepresentation ϕ_λ of $\mathcal{G}$. Note that this matrix is defined completely by C and the PTM representation of G. Note also that this matrix has only non-negative entries, that is $M_{\lambda,\hat{\lambda}} \geq 0 \quad \forall \lambda, \hat{\lambda}$.

In the following lemma we will assume that the mixing matrix M is not only non-negative but also irreducible in the Perron-Frobenius sense.[44] Formally this means that there exists an integer L such that A^L has only strictly positive entries. This assumption will allow us to invoke the powerful Perron-Frobenius theorem[44] to prove in Theorem 3 that interleaved character randomized benchmarking works as advertised. Below Theorem 3 we will also explicitly verify the irreducibility condition for 2-for-1 interleaved benchmarking with the CPHASE gate. We note that the assumption of irreducibility of M can be easily relaxed to M being a direct sum of irreducible matrices with the proof of Theorem 3 basically

unchanged. It is an open question if it can be relaxed further to encompass all non-negative mixing matrices.

Theorem 3. *Consider the outcome $k_{\lambda'}^m$ of an interleaved character randomized benchmarking experiment benchmarking group G, character group $\hat{G}$, subrepresentations $\hat{\phi} \subset \phi_{\lambda'}$ for some $\lambda' \in R_G$, interleaving gate C, and set of sequence lengths $\mathbb{M}$ and assume the existence of quantum channels $\mathcal{E}_C, \mathcal{E}$ s.t. $\tilde{C} = \mathcal{C} \mathcal{E}_C$ and $\tilde{\mathcal{G}} = \mathcal{E} \mathcal{G}$ for all $G \in G$. Now consider the matrix $M(\mathcal{E}_C \mathcal{E})$ as a function of the composed channel $\mathcal{E}_C \mathcal{E}$ with entries*

$$M_{\lambda,\hat{\lambda}}(\mathcal{E}_C \mathcal{E}) = \frac{1}{\text{Tr}(\mathcal{P}_\lambda)} \text{Tr}(\mathcal{P}_\lambda \mathcal{C} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger \mathcal{E}_C \mathcal{E}) \tag{56}$$

for $\lambda, \lambda' \in R'_G = R_G \backslash \{\text{id}\}$ where $\mathcal{P}_\lambda$ is again the projector onto the subrepresentation ϕ_λ of $\mathcal{G}$. If for $\mathcal{E} = \mathcal{E}_C = \mathcal{I}$ (the identity map) the matrix $M(\mathcal{I}) = M$ (the mixing matrix defined above) is irreducible (in the sense of Perron-Frobenius), then there exist parameters A, $f_{\lambda'}$ s.t.

$$|k_{\lambda'}^m - A f_{\lambda'}^m| \leq \delta_1 \delta_2^m \tag{57}$$

with $\delta_1 = O(1 - F_{\text{avg}}(\mathcal{E}_C \mathcal{E}))$ and $\delta_2 = \gamma + O([1 - F_{\text{avg}}(\mathcal{E}_C \mathcal{E})]^2)$ where γ is the second largest eigenvalue (in absolute value) of M. Moreover we have that (noting that $f_{\text{id}} = 1$ as the map $\mathcal{E}_C \mathcal{E}$ is CPTP):

$$\left| \frac{1}{2^q} \sum_{\lambda \in R_G} \text{Tr}(\mathcal{P}_\lambda) f_\lambda - \frac{2^q (F_{\text{avg}}(\mathcal{E}_C \mathcal{E}) + 1)}{2^q + 1} \right|$$

$$\leq O\left([1 - F_{\text{avg}}(\mathcal{E}_C \mathcal{E})]^2 \right) \tag{58}$$

Proof. Consider the definition of $k_{\lambda'}^m$:

$$k_m^{\lambda'} = |\hat{\phi}| \mathop{\mathbb{E}}_{\substack{\hat{G} \in \hat{G} \\ G_1, \dots, G_m \in G}} \chi_{\hat{\phi}}(\hat{G}) \langle\langle Q | \mathcal{E}_{\text{inv}} \mathcal{G}_{\text{inv}} \mathcal{C} \mathcal{E}_C \mathcal{E} \mathcal{G}_m$$

$$\times \mathcal{C} \mathcal{E}_C \mathcal{E} \dots \mathcal{C} \mathcal{E}_C \mathcal{E} \mathcal{G}_1 \hat{\mathcal{G}} | \rho \rangle\rangle, \tag{59}$$

where $G_{\text{inv}} = G_1^\dagger C^\dagger \cdots G_m^\dagger C^\dagger$ and $\mathcal{E}_{\text{inv}}$ is the noise associated to the inverse gate (which we assume to be constant). Using the character projection formula and Schur's lemma we can write this as

$$k_m^{\lambda'} = \mathop{\mathbb{E}}_{G_1, \dots, G_{m-1} \in G} \langle\langle Q | \mathcal{E}_{\text{inv}} \mathcal{G}_1^\dagger \mathcal{C}^\dagger \cdots \mathcal{G}_{m-1}^\dagger \mathcal{C}^\dagger$$

$$\times \left[\sum_{\lambda_m \in R'_G} \frac{\text{Tr}(P_{\lambda_m} \mathcal{E}_C \mathcal{E})}{\text{Tr}(\mathcal{P}_{\lambda_m})} \mathcal{P}_{\lambda_m} \right] \mathcal{C} \mathcal{E}_C \mathcal{E} \mathcal{G}_{m-1}$$

$$\times \mathcal{C} \mathcal{E}_C \mathcal{E} \dots \mathcal{C} \mathcal{E}_C \mathcal{E} \mathcal{G}_1 \mathcal{P}_{\hat{\phi}} | \rho \rangle\rangle. \tag{60}$$

Note now that in general $\mathcal{C}$ and $\mathcal{P}_{\lambda_m}$ do not commute. This means that we can not repeat the reasoning of Lemma 3 but must instead write (using Schur's lemma again):

$$k_m^{\lambda'} = \sum_{\lambda_m \in R'_G} \frac{\text{Tr}(P_{\lambda_m} \mathcal{E}_C \mathcal{E})}{\text{Tr}(\mathcal{P}_{\lambda_m})} \mathop{\mathbb{E}}_{G_1, \dots, G_{m-2} \in G} \langle\langle Q | \mathcal{E}_{\text{inv}} \mathcal{G}_1^\dagger \mathcal{C}^\dagger \cdots \mathcal{G}_{m-2}^\dagger \mathcal{C}^\dagger$$

$$\times \left[\sum_{\lambda_{m-1} \in R'_G} \frac{\text{Tr}(P_{\lambda_{m-1}} \mathcal{C}^\dagger P_{\lambda_m} \mathcal{C} \mathcal{E}_C \mathcal{E})}{\text{Tr}(\mathcal{P}_{\lambda_{m-1}})} \right]$$

$$\times \mathcal{P}_{\lambda_{m-1}} \mathcal{C} \mathcal{E}_C \mathcal{E} \mathcal{G}_{m-2} \mathcal{C} \mathcal{E}_C$$

$$\times \mathcal{E} \dots \mathcal{C} \mathcal{E}_C \mathcal{E} \mathcal{G}_1 \mathcal{P}_{\hat{\phi}} | \rho \rangle\rangle. \tag{61}$$

Here we recognize the definition of the matrix element $M_{\lambda_{m-1}, \lambda_m}(\mathcal{E}_C \mathcal{E})$. Moreover we can apply the above expansion to G_{m-2}, G_{m-3} and so forth writing the result in terms of powers of the matrix $M(\mathcal{E}_C \mathcal{E})$. After some reordering we get

$$k_m^{\lambda'} = \sum_{\lambda_1, \lambda_m \in R'_G} \frac{\text{Tr}(P_{\lambda_m} \mathcal{E}_C \mathcal{E})}{\text{Tr}(\mathcal{P}_{\lambda_m})} [M^{m-1}]_{\lambda_1, \lambda_m} \langle\langle Q | \mathcal{P}_{\lambda_1} \mathcal{P}_{\hat{\phi}} | \rho \rangle\rangle$$

where we have again absorbed the noise associated with the inverse G_{inv} into the measurement POVM element Q. Now recognizing that by

construction $\mathcal{P}_{\hat{\phi}} \subset \mathcal{P}_{\lambda'}$ we can write $k_m^{\lambda'}$ as

$$k_m^{\lambda'} = e_{\lambda'} M^m v^T \langle\langle Q|\mathcal{P}_{\hat{\phi}}|\rho\rangle\rangle \tag{62}$$

where $e_{\lambda'}$ is the λ'th standard basis row vector of length R_G' and $v = v(\mathcal{E}_C\mathcal{E})$ is a row vector of length R_G' with entries $[v]_\lambda = \frac{\text{Tr}(\mathcal{P}_\lambda \mathcal{E}_C\mathcal{E})}{\text{Tr}(\mathcal{P}_{\lambda m})}$. This looks somewhat like an exponential decay but not quite. Ideally we would like that M^m has one dominant eigenvalue and moreover that the vector v has high overlap with the corresponding eigenvector. This would guarantee that $k_m^{\lambda'}$ is close to a single exponential. The rest of the proof will argue that this is indeed the case. Now we use the assumption of the irreducibility of the mixing matrix $M = M(\mathcal{I})$. Subject to this assumption, the Perron-Frobenius theorem[44] states that the matrix M has a non-degenerate eigenvalue $\gamma_{\max}(M(\mathcal{I}))$ that is strictly larger in absolute value than all other eigenvalues of $M(\mathcal{I})$ and moreover satisfies the inequality

$$\min_{\lambda \in R_G'} \sum_{\hat{\lambda} \in R_G'} M_{\lambda,\hat{\lambda}} \leq \gamma_{\max}(M(\mathcal{I})) \leq \max_{\lambda \in R_G'} \sum_{\hat{\lambda} \in R_G'} M_{\lambda,\hat{\lambda}}. \tag{63}$$

It is easy to see from the definition of $M_{\lambda,\hat{\lambda}}$ that

$$\sum_{\hat{\lambda} \in R_G'} M_{\lambda,\hat{\lambda}} = \sum_{\hat{\lambda} \in R_G'} \frac{1}{\text{Tr}(\mathcal{P}_\lambda)} \text{Tr}(\mathcal{P}_\lambda \mathcal{C} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger) \tag{64}$$

$$= \sum_{\hat{\lambda} \in R_G'} \frac{1}{\text{Tr}(\mathcal{P}_\lambda)} \left(\mathcal{P}_\lambda \mathcal{C} \sum_{\hat{\lambda} \in R_G'} \mathcal{P}_{\hat{\lambda}} \mathcal{C}^\dagger \right) \tag{65}$$

$$= \frac{\text{Tr}(\mathcal{P}_\lambda)}{\text{Tr}(\mathcal{P}_\lambda)} = 1 \tag{66}$$

for all $\lambda \in R_G'$. This means the largest eigenvalue of $M(\mathcal{I})$ is exactly 1. Moreover, as one can easily deduce by direct calculation, the associated right-eigenvector is the vector $v^R = (1, 1, \ldots, 1)$. Note that this vector is precisely $v(\mathcal{E}_C\mathcal{E})$ (as defined in Eq. (62)) for $\mathcal{E}_C\mathcal{E} = \mathcal{I}$. Similarly the left-eigenvector of $M = M(\mathcal{I})$ is given by (in terms of its components) $v_\lambda^L = \text{Tr}(\mathcal{P}_\lambda)$. This allows us to calculate that $k_m^{\lambda'} = \langle\langle Q|\mathcal{P}_{\hat{\phi}}|\rho\rangle\rangle$ if $\mathcal{E}_C\mathcal{E} = \mathcal{I}$, which is as expected.

Now we will consider the map $\mathcal{E}_C\mathcal{E}$ as a perturbation of $\mathcal{I}$ with the perturbation parameter

$$\alpha = 1 - \frac{\text{Tr}(\mathcal{P}_{\text{tot}} \mathcal{E}_C\mathcal{E})}{\text{Tr}(\mathcal{P}_{\text{tot}})} \tag{67}$$

with $\mathcal{P}_{\text{tot}} = \sum_{\lambda \in R_G'} \mathcal{P}_\lambda$. We can write the quantum channel $\mathcal{E}_C\mathcal{E}$ as $\mathcal{E}_C\mathcal{E} = \mathcal{I} - \alpha\mathcal{F}$ where $\mathcal{F}$ is some superoperator (not CP, but by construction trace-annihilating). Since $M(\mathcal{E}_C\mathcal{E})$ is linear in its argument we can write $M(\mathcal{E}_C\mathcal{E}) = M(\mathcal{I}) - \alpha M(\mathcal{F})$. From standard matrix perturbation theory [ref. [45], Section 5.1] we can approximately calculate the largest eigenvalue of $M(\mathcal{E}_C\mathcal{E})$ as

$$\gamma_{\max}(M(\mathcal{E}_C\mathcal{E})) = \gamma_{\max}(M(\mathcal{I}))$$

$$- \alpha \frac{v^L M(\mathcal{F}) v^{R^T}}{v^L v^{R^T}} + O(\alpha^2) \tag{68}$$

We can now calculate the prefactor $\frac{v^L M(\mathcal{F}) v^{R^T}}{v^L v^{R^T}}$ as

$$\frac{v^L A(\mathcal{F}) v^{R^T}}{v^L v^{R^T}} = \frac{\sum_{\lambda \in R_G'} \sum_{\hat{\lambda} \in R_G'} v_\lambda^L M(\mathcal{F})_{\lambda,\hat{\lambda}} v_{\hat{\lambda}}^R}{\text{Tr}(\mathcal{P}_{\text{tot}})} \tag{69}$$

$$= \frac{\sum_{\lambda \in R_G'} \sum_{\hat{\lambda} \in R_G'} \text{Tr}(\mathcal{P}_\lambda \mathcal{C}^\dagger \mathcal{P}_{\hat{\lambda}} \mathcal{F})}{\text{Tr}(\mathcal{P}_{\text{tot}})} \tag{70}$$

$$= \frac{\text{Tr}(\mathcal{P}_{\text{tot}} \mathcal{F})}{\text{Tr}(\mathcal{P}_{\text{tot}})} \tag{71}$$

$$= \frac{1}{\alpha} \frac{\text{Tr}(\mathcal{P}_{\text{tot}}[\mathcal{I} - \mathcal{E}_C\mathcal{E}])}{\text{Tr}(\mathcal{P}_{\text{tot}})} \tag{72}$$

$$= 1 \tag{73}$$

where we used the definition of α in the last line. This means that $\gamma_{\max}(M(\mathcal{E}_C\mathcal{E})) = 1 - \alpha$ up to $O(\alpha^2)$ corrections. One could in principle calculate the prefactor of the correction term, but we will not pursue this here. Now we know that the matrix $M(\mathcal{E}_C\mathcal{E})^{m-1}$ in Eq. (62) will be dominated by a factor $(1 - \alpha + O(\alpha^2))^{m-1}$. However it could still be that the vector $v(\mathcal{E}_C\mathcal{E})$ in Eq. (62) has small overlap with the right-eigenvector

$v^R(\mathcal{E}_C\mathcal{E})$ of $M(\mathcal{E}_C\mathcal{E})$ associated to the largest eigenvalue $\gamma_{\max}(M(\mathcal{E}_C\mathcal{E}))$. We can again use a perturbation argument to see that this overlap will be big. Again from standard perturbation theory [ref. [45], Section 5.1] we have

$$\|v^R(\mathcal{E}_C\mathcal{E}) - v^R(\mathcal{I})\| = O(|\alpha|). \tag{74}$$

Moreover, by definition of $v^R(\mathcal{I})$ and $v(\mathcal{E}_C\mathcal{E})$ we have that $v^R v(\mathcal{E}_C\mathcal{E})^T = 1 - \alpha$. By the triangle inequality we thus have

$$\|v^R(\mathcal{E}_C\mathcal{E}) - v(\mathcal{E}_C\mathcal{E})\| = O(|\alpha|). \tag{75}$$

One can again fill in the constant factors here if one desires a more precise statement. Finally we note from Lemma 4 that

$$\alpha = 1 - \frac{\text{Tr}(\mathcal{P}_{\text{tot}} \mathcal{E}_C\mathcal{E})}{\text{Tr}(\mathcal{P}_{\text{tot}})} = \frac{2^q}{2^q - 1}(F(\mathcal{E}_C\mathcal{E}) - 1) \tag{76}$$

This means that in the relevant limit of high fidelity, α will be small, justifying our perturbative analysis. Defining γ to be the second largest (in absolute value) eigenvalue of $M(\mathcal{E}_C\mathcal{E})$, which by the same argument as above will be the second largest eigenvalue of $M(\mathcal{I})$ up to $O(\alpha^2)$ corrections, we get

$$|k_m^{\lambda'} - \langle\langle Q|\mathcal{P}_{\hat{\phi}}|\rho\rangle\rangle - \gamma_{\max}(M(\mathcal{E}_C\mathcal{E}))^{m-1} \langle\langle Q|\mathcal{P}_{\hat{\phi}}|\rho\rangle\rangle| \leq \delta_1 \delta_2^m$$

with $\delta_1 = O(1 - F_{\text{avg}}(\mathcal{E}_C\mathcal{E}))$ and $\delta_2 = |\gamma| + O((1 - F_{\text{avg}}(\mathcal{E}_C\mathcal{E}))^2)$. Moreover, we have from Eqs. (68) and (76) that

$$\gamma_{\max}(A(\mathcal{E}_C\mathcal{E})) = 1 - \frac{2^q}{2^q - 1}(F(\mathcal{E}_C\mathcal{E}) - 1)$$

$$+ O\left([1 - F_{\text{avg}}(\mathcal{E}_C\mathcal{E})]^2\right) \tag{77}$$

which immediately implies

$$\left| \frac{1}{2^q} \sum_{\lambda \in R_G} \text{Tr}(\mathcal{P}_\lambda) f_\lambda - \frac{2^q(F_{\text{avg}}(\mathcal{E}_C\mathcal{E}) + 1)}{2^q + 1} \right|$$

$$\leq O\left([1 - F_{\text{avg}}(\mathcal{E}_C\mathcal{E})]^2\right) \tag{78}$$

proving the lemma.

It is instructive to calculate the mixing matrix for a relevant example. We will calculate M for C the CPHASE gate and $G = C_1^{\otimes 2}$ two copies of the single qubit Clifford gates. Recall from the main text that the PTM representation of $C_1^{\otimes 2}$ has three non-trivial subrepresentations. From their definitions in Eq. (10) and the action of the CPHASE gate on the two qubit Pauli operators it is straightforward to see that the mixing matrix is of the form

$$M = \begin{pmatrix} 1/3 & 0 & 2/3 \\ 0 & 1/3 & 2/3 \\ 2/9 & 2/9 & 5/9 \end{pmatrix}. \tag{79}$$

Calculating M^2 one can see that M is indeed irreducible. Moreover M has eigenvalues 1, 1/3 and $-1/9$. This means that for 2-for-1 interleaved benchmarking the interleaved experiment produces data that deviates from a single exponential no more than $(1/3)^m$ (for sufficiently high fidelity) which will be negligible for even for fairly small m. This means that for 2-for-1 interleaved benchmarking the assumption that the interleaved experiment produces data described by a single exponential is good. We will see this confirmed numerically in the simulated experiment presented in Supplementary Fig. 2. Finally, we note that a similar result was achieved using different methods in ref. [46,47]

DATA AVAILABILITY

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

J.H., X.X., L.M.K.V. and S.W. conceived of the theoretical framework, detailed analysis was done by J.H. with input from X.X., L.M.K.V., and S.W., J.H. wrote the manuscript with input from X.X., L.M.K.V., and S.W., SW supervised the project.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Quantum Information* website (https://doi.org/10.1038/s41534-019-0182-7).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Dankert, C., Cleve, R., Emerson, J. & Livine, E. R. Exact and approximate unitary 2-designs: Constructions and applications. *Phys. Rev. A* **80**, 012304 (2006).
2. Magesan, E., Gambetta, J. M. & Emerson, J. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012).
3. Emerson, J., Alicki, R. & Życzkowski, K. Scalable noise estimation with random unitary operators. *J. Opt. B* **7**, S347 (2005).
4. Chow, J. M. et al. Randomized benchmarking and process tomography for gate errors in a solid-state qubit. *Phys. Rev. Lett.* **102**, 090502 (2009).
5. Gaebler, J. P. et al. Randomized benchmarking of multiqubit gates. *Phys. Rev. Lett.* **108**, 260503 (2012).
6. Granade, C., Ferrie, C. & Cory, D. G. Accelerated randomized benchmarking. *New J. Phys.* **17**, 013042 (2014).
7. Epstein, J. M., Cross, A. W., Magesan, E. & Gambetta, J. M. Investigating the limits of randomized benchmarking protocols. *Phys. Rev. A* **89**, 062321 (2014).
8. Knill, E. et al. Randomized benchmarking of quantum gates. *Phys. Rev. A* **77**, 012307 (2008).
9. Asaad, S. et al. Independent, extensible control of same-frequency super-conducting qubits by selective broadcasting. *npj Quantum Inf.* **2**, 16029 (2016).
10. Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
11. DiCarlo, L. et al. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240 (2009).
12. O'Malley, P. et al. Qubit metrology of ultralow phase noise using randomized benchmarking. *Phys. Rev. Applied* **3**, 044009 (2015).
13. Sheldon, S. et al. Characterizing errors on qubit operations via iterative randomized benchmarking. *Phys. Rev. A* **93**, 012301 (2016).
14. Wallman, J. J. Randomized benchmarking with gate-dependent noise. *Quantum* **2**, 47 (2018).
15. Proctor, T., Rudinger, K., Young, K., Sarovar, M. & Blume-Kohout, R. What randomized benchmarking actually measures. *Phys. Rev. Lett.* **119**, 130502 (2017).
16. Merkel, S. T., Pritchett, E. J. & Fong, B. H. Randomized benchmarking as convolution: Fourier analysis of gate dependent errors. *arXiv preprint arXiv:1804.05951* (2018).
17. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th edn (Cambridge University Press, New York, NY, USA, 2011).
18. Cross, A. W., Magesan, E., Bishop, L. S., Smolin, J. A. & Gambetta, J. M. Scalable randomised benchmarking of non-clifford gates. *npj Quantum Information* **2**, 16012 (2016).
19. Brown, W. G. & Eastin, B. Randomized benchmarking with restricted gate sets. *arXiv preprint arXiv:1801.04042* (2018).
20. Hashagen, A., Flammia, S., Gross, D. & Wallman, J. Real randomized benchmarking. *arXiv preprint arXiv:1801.06121* (2018).
21. França, D. S. & Hashagen, A.-L. Approximate randomized benchmarking for finite groups. *arXiv preprint arXiv:1803.03621* (2018).
22. Dugas, A. C., Wallman, J. J. & Emerson, J. Characterizing Universal Gate Sets via Dihedral Benchmarking. *arXiv preprint arXiv:1508.06312*.
23. Gambetta, J. M. et al. Characterization of addressability by simultaneous randomized benchmarking. *Phys. Rev. Lett.* **109**, 240504 (2012).
24. Harper, R. & Flammia, S. T. Estimating the fidelity of t gates using standard interleaved randomized benchmarking. *Quantum Sci.Technol.* **2**, 015008 (2017).
25. Flammia, S. T. & Liu, Y.-K. Direct fidelity estimation from few Pauli measurements. *Phys. Rev. Lett.* **106**, 230501 (2011).
26. Harper, R. & Flammia, S. T. Fault-tolerant logical gates in the ibm quantum experience. *Phys. Rev. Lett.* **122**, 080504 (2019).
27. Muhonen, J. T. et al. Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. *J. Phys. Condens. Matter* **27**, 154205 (2015).
28. Helsen, J., Wallman, J. J., Flammia, S. T. & Wehner, S. Multi-qubit randomized benchmarking using few samples. *arXiv preprint arXiv:1701.04299* (2017).
29. Fulton, W. & Harris, J. *Representation Theory: A First Course*. Readings in Mathematics (Springer-Verlag, New York, 2004).
30. Wallman, J. J. & Flammia, S. T. Randomized benchmarking with confidence. *New J. Phys.* **16**, 103032 (2014).
31. Wolf, M. Quantum channels operations: Guided tour. *Lecture Notes* (2012). http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines///MichaelWolf/QChannelLecture.pdf.
32. Chuang, I. L. & Nielsen, M. A. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* **44**, 2455 (1997).
33. Xue, X. et al. Benchmarking gate fidelities in a si/sige two-qubit device. *arXiv preprint arXiv:1811.04002* (2018).
34. Magesan, E. et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.* **109**, 080505 (2012).
35. Dugas, A. C., Wallman, J. J. & Emerson, J. Efficiently characterizing the total error in quantum circuits. *arXiv preprint arXiv:1610.05296* (2016).
36. Watson, T. et al. A programmable two-qubit quantum processor in silicon. *Nature* **555**, 633 (2018).
37. Zajac, D. M. et al. Resonantly driven cnot gate for electron spins. *Science* **359**, 439–442 (2018).
38. Huang, W. et al. Fidelity benchmarks for two-qubit gates in silicon. *arXiv preprint arXiv:1805.05027* (2018).
39. Kalra, R., Laucht, A., Hill, C. D. & Morello, A. Robust two-qubit gates for donors in silicon controlled by hyperfine interactions. *Physical Review X* **4**, 021044 (2014).
40. Li, R. et al. A crossbar network for silicon quantum dot qubits. *Science advances* **4**, eaar3960 (2018).
41. Hincks, I., Wallman, J. J., Ferrie, C., Granade, C. & Cory, D. G. Bayesian inference for randomized benchmarking protocols. *arXiv preprint arXiv:1802.00401* (2018).
42. Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journ. Am. Stat. Assoc.* **58**, 13–30 (1963).
43. Watrous, J. Notes on super-operator norms induced by schatten norms. *arXiv preprint arXiv:0411077* (2004).
44. MacCluer, C. R. The many proofs and applications of Perron's theorem. *Siam Review* **42**, 487–498 (2000).
45. Sakurai, J. J. et al. *Modern quantum mechanics*, vol. 261 (Pearson, 2014).
46. Erhard, A. et al. Characterizing large-scale quantum computers via cycle benchmarking. *arXiv preprint arXiv:1902.08543* (2019).
47. Wallman, J. J. & Emerson, J. Determining the capacity of any quantum computer to perform a quantum computation. *In Preparation* (2018).