A new class of efficient randomized benchmarking protocols

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

[1] *QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*
[2] *Kavli Institute of Nanoscience, Delft University of Technology, 2600 GA Delft, The Netherlands*

(Dated: June 7, 2018)

Randomized benchmarking is a technique for estimating the average fidelity of a set of quantum gates. For general gatesets, however, it is difficult to draw robust conclusions from the resulting data. Here we propose a new method based on representation theory that has little experimental overhead and applies to a broad class of benchmarking problems. As an example, we apply our method to a gateset that includes the T-gate, and analyze a new interleaved benchmarking protocol that extracts the average fidelity of a 2-qubit Clifford gate using only single-qubit Clifford gates as reference.

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]. Keys to the wide adoption of randomized benchmarking are its scalability with respect to the number of qubits in the quantum device, and its insensitivity to errors in state preparation and measurement. It has also recently been shown to be relatively insensitive to variations in the error associated to different gates in the implementation [14–16].

The randomized benchmarking protocol is defined with respect to a 'gateset', a discrete collection of quantum gates that usually has some special structure. Usually the gateset is assumed to have a *group* structure, an assumption we will make throughout this Letter. When implemented, the protocol returns quality parameters associated to the entire gateset, often interpreted as the average quality parameter of this gateset. The most important example of such a quality parameter is the average fidelity which can always be obtained as a weighted average of the quality parameters yielded by randomized benchmarking. We will now give an overview of the randomized benchmarking procedure for a gateset $\mathbb{G}$ with initial state ρ and a binary measurement represented by a two-component POVM $\{Q, \mathbb{1}-Q\}$.

Randomized benchmarking is performed by sampling sequences of a fixed length m of gates from the gateset $\mathbb{G}$. These sequences are applied to an initial state, followed by a global inversion gate such that in the ideal case the system is returned to the starting state. Then the overlap between the output state and the initial state is estimated. This is repeated for many sequences of the same length and averaged yielding a single average probability p_m. Repeating this procedure for various sequence lengths yields a set of averages $\{p_m\}_m$.

What to do next depends on the gateset $\mathbb{G}$. When $\mathbb{G}$ is the Clifford group [17] it can be shown (under the assumption of gate-independent noise) that the data-points $\{p_m\}_m$ can be fitted to a single exponential decay of the form

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

where A, B only depend on how well the state ρ was prepared and measured and the quality parameter f only depends on how well the gates in $\mathbb{G}$ are implemented. This parameter f can then be straightforwardly related to a standard metric of gate quality known as the average fidelity F_{avg} [2]. This fitting relation holds intuitively because averaging over all elements of the Clifford group causes the noise process affecting the gates to be 'smeared out', in the sense that it affects any possible input state in the same manner. This averaged noise then accretes exponentially with sequence length m.

However when the gateset is not the multi-qubit Clifford group but rather some other group [18–23] or even when the noise affecting the Clifford group gates is not trace-preserving [24] the fitting relation in eq. (1) does not hold. Instead it has to be generalized to

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

where the f_λ are general quality parameters that only depend on the gates being implemented and the A_λ prefactors depend only on how well the state ρ is prepared and measured. Intuitively this is so because the averaging process does not smear out the noise into affecting all states equally. Rather the state space will split into several 'sectors' labeled by λ (in a sense described more formally below) such that states within the same sector experience the same noise but the noise varies from sector to sector. The average fidelity associated to the gateset can then always be found as a weighted average of the quality parameters f_λ.

This multi-parameter fitting problem will be difficult to perform in practice. One could try to fit the randomized benchmarking data $\{p_m\}_m$ to a linear combination of exponential decays directly, but this will generally

result in poor confidence intervals around the fitting parameters unless large amounts of data are gathered. More fundamentally, it will then be impossible to relate the various decay factors to the correct decay channel in eq. (2), making it impossible to reliably estimate the average fidelity of the gateset, even in the limit of infinite data.

In current literature [19, 20] this issue is sidestepped by performing the experiment 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], (2) it is, for more general gatesets, not always possible to efficiently find such a maximizing state ρ_λ and (3) both previous problems become more pronounced as the number of quality parameters f_λ increases, limiting practical randomized benchmarking to gatesets that generate only a small amount of quality parameters.

In this Letter we propose an adaptation of the randomized benchmarking procedure, which we call character randomized benchmarking, which solves the above problems. 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, 26, 27], can be regarded as special cases of our method.

For the rest of this Letter we will work in the Pauli Transfer Matrix (PTM) picture [28]. This means we think of states ρ and POVM elements Q as vectors and co-vectors which we denote $|\rho\rangle$ and $\langle Q|$ respectively. Quantum channels $\mathcal{E}$ [29] are then matrices (we will denote a channel and its PTM representation by the same letter) and we have $\mathcal{E}|\rho\rangle = |\mathcal{E}(\rho)\rangle$. Composition of channels also corresponds to multiplication of their PTM representations, that is $|\mathcal{E}\circ\mathcal{F}(\rho)\rangle = \mathcal{E}\mathcal{F}|\rho\rangle$. Moreover we can write expectation values as bra-ket inner products, i.e. $\langle Q|\mathcal{E}|\rho\rangle = \mathrm{Tr}(Q\mathcal{E}(\rho))$. For a unitary gate G we will denote its ideal action on ρ as $\mathcal{G}$, i.e. $\mathcal{G}|\rho\rangle = |G\rho G^\dagger\rangle$ and we will denote its real (noisy) implementation by $\widetilde{\mathcal{G}}$. For a general randomized benchmarking procedure over a group $\mathbb{G}$ (with input state ρ and measurement POVM $\{Q, \mathbb{1}-Q\}$) we can rewrite [30] the exact average over all sequences of length m as

$$p_m = \langle Q| \left(\frac{1}{|\mathbb{G}|} \sum_{G\in\mathbb{G}} \mathcal{G}^\dagger \widetilde{\mathcal{G}} \right)^m |\rho\rangle. \tag{3}$$

The key insight to randomized benchmarking, which formalizes the intuition of 'smearing' mentioned above, is that $\mathcal{G}$ is a *representation* of $G \in \mathbb{G}$. This representation will not in general be irreducible but will rather decompose into irreducible subrepresentations, that is $\mathcal{G} = \bigoplus_\lambda \phi_\lambda(G)^{\otimes m_\lambda}$ where ϕ_λ is an irreducible subrepresentation of $\mathbb{G}$ that appears with multiplicity m_λ (that is, there are m_λ equivalent copies of ϕ_λ present in $\mathcal{G}$). These irreducible subrepresentations formalize the notion of 'sectors' mentioned above. For simplicity we will here assume that all irreducible subrepresentations ϕ_λ have multiplicity $m_\lambda = 1$, however our results will remain valid in the presence of multiplicities (see Supplementary Material). Using Schur's lemma, a fundamental result in representation theory, we can then write eq. (3) as

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

where $\mathcal{P}_\lambda$ is the projector onto the support of ϕ_λ and f_λ is the quality factor associated to ϕ_λ. This reproduces eq. (2). For more details on representation theory see [31] or the Supplementary Materials. A formal proof of eq. (4) can e.g. be found in [21]. Associated to any representation ϕ of a group $\mathbb{G}$ is a character function $\chi : \mathbb{G} \to \mathbb{R}$, from the group to the real numbers [32]. Associated to this function is the following projection formula

$$\frac{1}{|\mathbb{G}|} \sum_{G\in\mathbb{G}} \chi(G)\phi(G) = \frac{1}{\mathrm{Tr}(\mathcal{P})}\mathcal{P}, \tag{5}$$

where $\mathcal{P}$ is the projector onto the support of ϕ. We here 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). To see this consider a group $\mathbb{G}$ and a parameter $f_{\lambda'}$ associated to a specific subrepresentation $\phi_{\lambda'}$ of $\mathcal{G}$ for $G \in \mathbb{G}$. Now consider a group $\hat{\mathbb{G}}$ (we might choose $\hat{\mathbb{G}} = \mathbb{G}$ or $\hat{\mathbb{G}} \subset \mathbb{G}$) such that $\hat{\mathcal{G}} = \oplus_{\hat\lambda}\phi_{\hat\lambda}(\hat G)$ for $\hat G \in \hat{\mathbb{G}}$ and in particular such that there exists a $\hat\lambda'$ such that its associated representation $\phi_{\hat\lambda'}$ has support inside the representation $\phi_{\lambda'}$ of $\mathbb{G}$. The representation $\phi_{\hat\lambda'}$ has character function $\chi_{\hat\lambda'}$. Now we can consider the following adapted randomized benchmarking protocol which we call character randomized benchmarking.

1. Sample $\vec G = G_1, \ldots, G_m$ uniformly at random from $\mathbb{G}$
2. Sample $\hat G$ uniformly at random from $\hat{\mathbb{G}}$
3. Prepare a quantum state ρ and apply the gates $(G_1\hat G), G_2, \ldots G_m$
4. Compute the inverse $G_{\mathrm{inv}} = (G_m \cdots G_1)^\dagger$ and apply it (note that $\hat G$ is not inverted)
5. Estimate the weighted 'survival probability' $k_m^{\hat\lambda'}(\vec G, \hat G) = \mathrm{Tr}(P_{\hat\lambda'})\chi_{\hat\lambda'}(\hat G)\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}} \times\widetilde{\mathcal{G}}_m \cdots \widetilde{(\mathcal{G}_1\hat{\mathcal{G}})}|\rho\rangle$
6. Repeat for many $\hat G \in \hat{\mathbb{G}}$ and estimate the average $k_m^{\hat\lambda'}(\vec G) = \mathbb{E}_{\hat G}(k_m^{\hat\lambda'}(\vec G, \hat G))$

7. Repeat for many $\vec{G}$ and estimate the average $k_m = \mathbb{E}_{\vec{G}}(k_m^{\hat{\lambda}'}(\vec{G}))$
8. Repeat for many different m

The major difference between the standard and character randomized benchmarking protocols is the introduction of an extra average over a group $\hat{\mathbb{G}}$. This extra gate $\hat{G} \in \hat{\mathbb{G}}$ is not included when computing the global inverse G_{inv}. Note that this extra gate can be compiled into the sequence of gates $(G_1, \ldots, G_m)$ and thus does not result in extra noise. The average over the elements of $\hat{\mathbb{G}}$ is also weighted by the character function $\chi_{\hat{\lambda}'}$ associated to the representation $\phi_{\hat{\lambda}'}$. Similar to eq. (3) we can rewrite the uniform average over all $\vec{G} \in \mathbb{G}^{\times m}$ and $\hat{G} \in \hat{\mathbb{G}}$ as

$$k_m = \frac{\text{Tr}(\mathcal{P}_{\hat{\lambda}'})}{|\hat{\mathbb{G}}|}\langle Q|\left[\frac{1}{|\mathbb{G}|}\sum_{G\in\mathbb{G}} \mathcal{G}^\dagger \widetilde{\mathcal{G}}\right]^m \sum_{\hat{G}\in\hat{\mathbb{G}}} \chi_{\hat{\lambda}'}(\hat{G})\hat{\mathcal{G}}|\rho\rangle.$$

Using the character projection formula (eq. (5)) and the standard randomized benchmarking representation theory formula (eq. (4)) we can write this as

$$k_m = \sum_{\lambda} \langle Q|\mathcal{P}_\lambda \mathcal{P}_{\hat{\lambda}'}|\rho\rangle f_\lambda^m = \langle Q|\mathcal{P}_{\hat{\lambda}'}|\rho\rangle f_{\lambda'}^m \qquad (6)$$

since we have chosen $\mathbb{G}$ and $\hat{\lambda}'$ such that the support of $\mathcal{P}_{\hat{\lambda}'}$ is a subspace of the support of $\mathcal{P}_{\lambda'}$. This means the character randomized benchmarking protocol isolates the exponential decay associated to the quality factor $f_{\lambda'}$ independent of state preparation and measurement, although of course we would like to choose Q and ρ in a way that maximizes $\langle Q|\mathcal{P}_{\hat{\lambda}'}|\rho\rangle$. Repeating this procedure for different choices of λ' and $\hat{\lambda}'$ we can reliably reconstruct all quality parameters associated with randomized benchmarking over the group $\mathbb{G}$, which can then always be combined into the average fidelity of the group $\mathbb{G}$.

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 [20] while the second one, performing interleaved benchmarking on a 2-qubit gate using only single qubit gates a reference, is a new protocol.

Benchmarking T-gates. The most common universal gateset considered in the literature is the Clifford+T gateset. The average fidelity of the Clifford gates can be extracted using standard RB over the Clifford group but to investigate the average fidelity of the T gate a different approach is needed. One choice is to perform randomized benchmarking over the group generated by the CNOT, Pauli X and T gates. This group is an example of a CNOT-dihedral group and its use for randomized benchmarking was investigated in [20]. There it was derived that randomized benchmarking over this group leads to a fit involving 3 quality parameters f_1, f_2, f_3 and hence a fitting problem involving 3 exponential decays. The average fidelity of the gates can then be derived as a weighted average of the quality parameters f_1, f_2, f_3.

Since there are 3 decay channels the PTM representation of this group has 3 irreducible subrepresentations labeled ϕ_1, ϕ_2, ϕ_3. The associated projectors for ϕ_2, ϕ_3 are given in in the PTM picture as

$$\mathcal{P}_2 = 2^{-n}\sum_{P\in\mathcal{Z}/\{\mathbb{1}\}} |P\rangle\langle P|, \quad \mathcal{P}_3 = 2^{-n}\sum_{P\in\mathbb{P}/\mathcal{Z}} |P\rangle\langle P|,$$

where n is the number of qubits in the system, $\mathbb{P}$ is the Pauli group and $\mathcal{Z}$ is the subgroup of the Pauli group composed only of tensor products of Z and $\mathbb{1}$. Let's say we want to isolate the quality factor f_2 (associated to ϕ_2). We must begin by choosing a group $\hat{\mathbb{G}}$. A good choice for $\hat{\mathbb{G}}$ would be the Pauli group $\mathbb{P}$. Note that $\mathbb{P} \subset \mathbb{G}$.

The PTM representation of the Pauli group has 2^n irreducible inequivalent subrepresentations associated to each basis element Pauli matrix P'. The character associated to this representation will be $\chi_{P'}(P) = (-1)^{\langle P,P'\rangle}$ where $\langle P, P'\rangle = 1$ if and only if P and P' anti-commute and zero otherwise. Note that every choice of P' leads to a different character function $\chi_{P'}$. Hence to isolate the the exponential decay f_2^m associated to ϕ_2 we have to choose $P' \subset \mathcal{Z}$. The character randomized benchmarking experiment can then be described by

$$k_m = 2^{-n}\langle Q|P'\rangle\langle P'|\rho\rangle f_2^m \qquad (7)$$

hence the optimal input state for this experiment would be $\rho = 2^{-n}(\mathbb{1} + P')$, which is in general not pure but can be prepared by performing the experiment multiple times for pure states in the support of ρ and then averaging over these states. Note that one only needs to sample a constant number (independent of the number of qubits) of of these states to compute any expectation value involving ρ (see Supplementary Materials for details). Note also that the factor 2^{-n} is just a normalization factor and that if ρ is prepared as above the signal k_m does not go down exponentially with the number of qubits (but rather stays constant).

Benchmarking a 2-qubit gate using only single qubit gates. The next example is a new protocol. It is a way to perform interleaved randomized benchmarking of a 2-qubit Clifford gate using only single qubit gates as reference gates. The advantages of this are (1) lower experimental requirements and (2) high fidelity of the reference gates relative to the interleaved gate which allows for a tighter bound on the average fidelity of the interleaved gate using the techniques of [33].This is assuming single qubit gates have higher fidelity than two qubit gates, which is the case in most quantum computing platforms. An interleaved benchmarking experiment consists of two stages, a reference experiment and

an interleaved experiment. The reference experiment will in this case be character randomized benchmarking using 2 copies of the single-qubit Clifford group $\mathbb{G} = \mathbb{C}_1^{\otimes 2}$ (this is also the group considered in [19]). The fitting curve 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 (in the PTM picture) onto the associated irreducible representations ϕ_w are

$$\mathcal{P}_w = 2^{-2} \sum_{P \in \mathbb{P}_w} |P\rangle\langle P| \tag{8}$$

where $\mathbb{P}_w$ is the subset of the 2-qubit Pauli group, the elements of which are composed of non-identity Pauli matrices at the i'th tensor factor if and only if $w_i = 1$. The average fidelity F_{avg} can be obtained from the quality parameters as $F_{\text{avg}} = (3/15)(f_{01} + f_{10}) + (9/15)f_{11}$ (see Supplementary Material). We can isolate the exponential decay associated to each quality parameter f_w by again choosing $\hat{\mathbb{G}} = \mathbb{P}$ the (2-qubit) Pauli group and for each w running the character randomized benchmarking protocol using again the character function $\chi_{P'}$ of a representation carried by the Pauli basis vector $|P'\rangle$ where for each w the choice of P' is such that $P' \in \mathbb{P}_w$. Once we have obtained all relevant quality parameters f_w we can compute the average reference fidelity F_{ref}. Next we repeat the character randomized benchmarking experiment but for every sequence $\vec{G} = (G_1, \ldots, G_m)$ we apply the sequence $(G_1, C, G_2, \ldots, C, G_m)$ 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. This interleaved experiment allows us to compute the interleaved fidelity F_{int}. From F_{ref} and F_{int} we can then compute upper and lower bounds on the average fidelity $F_{\text{avg}}(C)$ of the gate C in particular. This is detailed in [33].

Scalability of the protocol. The character randomized benchmarking protocol presented above is not necessarily scalable in system size. The first problem is that, for an arbitrary group $\hat{\mathbb{G}}$ the representation one might want to project onto could have a dimension that grows exponentially in the number of qubits in the system. Similarly the character function could have values that grow exponentially in the number of qubits. This means the quantity $k_m(\vec{G})$ can not be reliably estimated for experiments involving more than a few qubits. A solution to this is to choose $\hat{\mathbb{G}}$ such that the representation being projected onto has small dimension. Since the maximal absolute value of the character function is bounded by the dimension of the associated representation [31] the value of the character function will also be small. This was the case in the two examples presented above where we chose $\hat{\mathbb{G}} = \mathbb{P}$ which had only one-dimensional subrepresentations. There is some trade-off to be made here since choosing a group with smaller irreducible subrepresentations can make it harder to prepare an input state that generates sufficiently high signal and also one might run into situations where $\hat{\mathcal{G}}$ contains multiple copies of the same representation, which complicates matters (see the Supplementary Materials). However, when benchmarking any group $\mathbb{G}$ which has the Pauli group $\mathbb{P}$ as a subgroup one can always set $\mathbb{P} = \hat{\mathbb{G}}$ and project onto one of the one-dimensional subrepresentations as described above.

Finite Sampling. The character randomized benchmarking protocol involves averages over sets that are exponentially large in the number of qubits and the sequence length m. One can therefore only estimate these averages empirically by sampling a small subset of these sets. The quality of these empirical estimates can then be ascertained using confidence intervals. There are two different averages present in character randomized benchmarking. Firstly the character-weighted average over the group $\hat{\mathbb{G}}$ for a single sequence $\vec{G}$ (denoted $k_m^{\hat{\lambda}'}(\vec{G})$) and secondly the average over sequences $\vec{G} \in \mathbb{G}^{\times m}$ (denoted k_m). These have to be treated differently.

The character-weighted average over the group $\hat{\mathbb{G}}$ for a single sequence $\vec{G}$, $k_m^{\lambda'}(\vec{G})$ can be estimated by not estimating each character-weighted expectation value $k_m^{\hat{\lambda}'}(\vec{G}, \hat{G})$ individually but rather estimate $\vec{G}$ $k_m^{\hat{\lambda}'}(\vec{G})$ directly by the following procedure

1. Sample $\hat{G} \in \hat{\mathbb{G}}$ uniformly at random
2. Prepare the state $\mathcal{G}_{\text{inv}}\mathcal{G}_m \cdots \mathcal{G}_1 \hat{\mathcal{G}}|\rho\rangle$ and measure it once obtaining a result $b(\hat{G}) \in \{0,1\}$
3. Compute
   $x(\hat{G}) = \chi_{\hat{\lambda}'}(\hat{G}) \operatorname{Tr}(\mathcal{P}_{\hat{\lambda}'}) b(\hat{G}) \in \{0, \chi_{\hat{\lambda}'}(\hat{G}) \operatorname{Tr}(\mathcal{P}_{\hat{\lambda}'})\}$
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 $k_m^{\hat{\lambda}'}(\vec{G})$ that takes values in the interval $[-\chi^*_{\hat{\lambda}'}, \chi^*_{\hat{\lambda}'}]$ where $\chi^*_{\hat{\lambda}'}$ is the largest value of the character function $\chi_{\hat{\lambda}'}$. For the examples given above this maximal value is 1. Using standard statistical techniques [34] we can give e.g. a 99% confidence interval of size 0.02 around $k_m^{\hat{\lambda}'}(\vec{G})$ 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.

In [27, 35] it was shown that the average k_m over sequences $\vec{G} \in \mathbb{G}^{\times m}$ can be estimated with high precision and high confidence using only a few hundred sequences. These results, which were derived for standard randomized benchmarking, can be straight-forwardly extended

to character randomized benchmarking. Whether similar results when performing benchmarking with other groups is however an open question. This issue is not unique to character randomized benchmarking but is equally present when performing standard randomized benchmarking over any non-Clifford group.

We would like to thank Thomas Watson, Jérémy Ribeiro and Bas Dirkse for enlightening discussions. JH and SW are funded by STW Netherlands, NWO VIDI, an ERC Starting Grant and by the NWO Zwaartekracht QSC grant. XX and LMKV are funded by the Army Research Office (ARO) under Grant Number W911NF-17-1-0274.

[1] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77** (2008), 10.1103/physreva.77.012307.
[2] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85** (2012), 10.1103/physreva.85.042311.
[3] J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B **7**, S347 (2005).
[4] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102**, 090502 (2009).
[5] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. Lett. **108**, 260503 (2012).
[6] C. Granade, C. Ferrie, and D. G. Cory, New J. Phys. **17**, 013042 (2014), arXiv:1404.5275v1.
[7] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, Phys. Rev. A **89**, 062321 (2014), arXiv:1308.2928 [quant-ph].
[8] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).
[9] S. Asaad, C. Dickel, N. K. Langford, S. Poletto, A. Bruno, M. A. Rol, D. Deurloo, and L. DiCarlo, npj Quantum Inf. **2**, 16029 (2016), arXiv:1508.06676.
[10] R. Barends, J. Kelly, A. Megrant, D. Veitia, A.Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. J. J. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Nature **508**, 500 (2014).
[11] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Nature **460**, 240 (2009), arXiv:0903.2030.
[12] P. OMalley, J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, I.-C. Hoi, *et al.*, Phys. Rev. Applied **3**, 044009 (2015).
[13] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Phys. Rev. A **93**, 012301 (2016).
[14] J. J. Wallman, Quantum **2**, 47 (2018).
[15] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Phys. Rev. A **93**, 022303 (2016).
[16] S. T. Merkel, E. J. Pritchett, and B. H. Fong, arXiv preprint arXiv:1804.05951 (2018).
[17] Or more generally any group that is a 2-design [30].
[18] A. Hashagen, S. Flammia, D. Gross, and J. Wallman, arXiv preprint arXiv:1801.06121 (2018).
[19] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109** (2012).
[20] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, npj Quantum Information **2** (2016).
[21] D. S. França and A.-L. Hashagen, arXiv preprint arXiv:1803.03621 (2018).
[22] A. C. Dugas, J. J. Wallman, and J. Emerson, (2015), arXiv:1508.06312.
[23] W. G. Brown and B. Eastin, arXiv preprint arXiv:1801.04042 (2018).
[24] In the experimental community this type of noise is often referred to as leakage.
[25] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).
[26] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, Journal of Physics Condensed Matter **27**, 154205 (2015), arXiv:1410.2338 [quant-ph].
[27] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, arXiv preprint arXiv:1701.04299 (2017).
[28] This picture is also sometimes called the Liouville representation or affine representation of quantum channels. [35, 36].
[29] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).
[30] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
[31] W. Fulton and J. Harris, *Representation Theory: A First Course*, Readings in Mathematics (Springer-Verlag New York, 2004).
[32] Generally the character function is a map to the complex numbers, but in our case it is enough to only consider real representations.
[33] 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, Phys. Rev. Lett. **109** (2012).
[34] W. Hoeffding, Journ. Am. Stat. Assoc. **58**, 13 (1963).
[35] J. J. Wallman and S. T. Flammia, New J. Phys. **16**, 103032 (2014).
[36] M. Wolf, Lecture Notes (2012).
[37] R. Goodman and N. R. Wallach, *Symmetry, Representations, and Invariants*, Graduate Texts in Mathematics (Springer, 2009).
[38] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).
[39] M. B. Ruskai, S. Szarek, and E. Werner, Lin. Alg. and its Appl. **347**, 159 (2002).
[40] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).
[41] J. Helsen, J. J. Wallman, and S. Wehner,

arXiv:1609.08188, to appear in Journ. Math. Phys. (2016).
[42] D. M. Titterington, A. F. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions* (Wiley,, 1985).

APPENDIX A: BACKGROUND MATERIAL

In this section we present, for the benefit of the reader, some well known facts about representation theory and the representation of quantum channels. In particular we will review representations and characters and explain in more detail the Pauli transfer matrix formalism for quantum channels. We also treat the representations of the Pauli group in more detail. More background on representations and characters can be found in [31, 37] while our presentation of quantum channels is based on [36, 38].

Representation theory

We recall some useful facts about the representations of finite groups. For a more in depth treatment of this topic we refer to [31, 37]. Let $\mathbb{G}$ be a finite group and let V be some finite dimensional complex vector space. Let also $GL(V)$ be the group of invertible linear transformations of V. We can define a *representation* ϕ of the group G on the space V as a map

$$\phi : \mathbb{G} \to \mathrm{GL}(V) : G \mapsto \phi(G) \tag{9}$$

that has the property

$$\phi(G)\phi(H) = \phi(GH), \qquad \forall G, H \in \mathbb{G}. \tag{10}$$

In general we will assume the operators $\phi(G)$ to be unitary. If there is a non-trivial subspace W of V such that for all vectors $w \in W$ we have

$$\phi(G)w \in W, \qquad \forall G \in \mathbb{G}, \tag{11}$$

then the representation ϕ is called *reducible*. The restriction of ϕ to the subspace W is also a representation, which we call a *subrepresentation* of ϕ. If there are no non-trivial subspaces W such that eq. (11) holds the representation ϕ is called *irreducible*. Two representations ϕ, ϕ' of a group $\mathbb{G}$ on spaces V, V' are called *equivalent* if there exists an invertible linear map $T : V \to V'$ such that

$$T \circ \phi(G) = \phi'(G) \circ T, \qquad \forall G \in \mathbb{G}. \tag{12}$$

We will denote this by $\phi \equiv \phi'$. For a representation ϕ on a vector space V we can, for all linear maps $A : V \to V$ also define the *twirl* of A with respect to ϕ. This is denoted as $\mathcal{T}$ and has the form

$$\mathcal{T}_\phi(A) := \frac{1}{|\mathbb{G}|} \sum_{G \in \mathbb{G}} \phi(G) A \phi(G)^\dagger. \tag{13}$$

The following corollary of Schur's lemma, an essential result from representation theory [31, 37], allows us to evaluate twirls over certain types of representations.

Lemma 1 (Lemma 1.7 and Prop. 1.8 in [31])**.** *Let $\mathbb{G}$ be a finite group and let ϕ be a representation of $\mathbb{G}$ on a complex vector space V with decomposition*

$$\phi(G) \simeq \bigoplus_i \phi_i(G)^{\otimes m_i}, \qquad \forall G \in \mathbb{G} \tag{14}$$

into inequivalent irreducible subrepresentations ϕ_i where m_i is the multiplicity of ϕ_i. Then for any linear map $A : V \to V$ the twirl of A over G takes the form

$$\mathcal{T}_\phi(A) = \frac{1}{|\mathbb{G}|} \sum_{G \in \mathbb{G}} \phi(g) A \phi(g)^\dagger = \sum_i \sum_{j,j' \in \{1,\ldots,m_i\}} \frac{\mathrm{Tr}(A P_{j_i}^{j'_i})}{\mathrm{Tr}(P_{j_i}^{j_i})} P_{j_i}^{j'_i}. \tag{15}$$

where $P_{j_i}^{j'_i}$ is a linear map from the support of the j_i'th copy of ϕ_i to the support of the j'_i'th copy of ϕ_i. Note that $P_{j_i}^{j_i}$ is a projector onto the support of the j_i'th copy of ϕ_i.

When the decomposition of ϕ has no multiplicities, i.e.

$$\phi(G) \simeq \bigoplus_i \phi_i(G), \qquad \forall G \in \mathbb{G} \tag{16}$$

where all ϕ_i are irreducible inequivalent representations then the formula in lemma 1 becomes significantly simpler. We have for all linear maps $A : V \to V$

$$\mathcal{T}_\phi(A) = \sum_i \frac{\mathrm{Tr}(AP_i)}{\mathrm{Tr}(P_i)} P_i, \tag{17}$$

where P_i is the projector onto the support of the irreducible subrepresentation ϕ_i.

We next recall the character of a representation. Let $\phi : \mathbb{G} \to V$ be a representation of a finite group $\mathbb{G}$ on a finite dimensional (real or complex) vector space V. The character $\chi_\phi : \mathbb{G} \to \mathbb{R}$ of the representation ϕ is defined as

$$\chi_\phi : \mathbb{G} \to \mathbb{R} : G \mapsto \chi_\phi(G) = \mathrm{Tr}_V(\phi(G)), \tag{18}$$

where $\mathrm{Tr}_V(\ )$ denotes the trace over the vector space V. Note that characters can in general be complex (that is, functions from $\mathbb{G}$ to $\mathbb{C}$) but we will only consider representations with real character here. Characters have a number of useful properties [31] which we recall here. For representations φ, φ' we have the relations

$$\chi_{\varphi\otimes\varphi'} = \chi_\phi \, \chi_{\phi'}, \tag{19}$$

$$\chi_{\varphi\oplus\varphi'} = \chi_\phi + \chi_{\phi'}, \tag{20}$$

with suitable generalizations to multiple direct sums and tensor products. The following lemma, often referred to as the generalized projection formula, is of great use to us

Lemma 2 (Formula 2.32 in [31])**.** *Let $\mathbb{G}$ be a group and let ϕ be a representation of $\mathbb{G}$. Let also ϕ' be an irreducible subrepresentation of ϕ with associated character function $\chi_{\phi'}$. Then the following formula holds*

$$\frac{\mathrm{Tr}(\mathcal{P}')}{|\mathbb{G}|} \sum_{G\in\mathbb{G}} \chi_{\phi'}(G)\phi(G) = \sum_{\hat{\phi}\equiv\phi'} P_{\hat{\phi}}, \tag{21}$$

where $P_{\hat{\phi}}$ is the projector onto the support of $\hat{\phi}$ which is a subrepresentation of ϕ (equivalent to ϕ'). Note also that the sum runs over all representations $\hat{\phi}$ in ϕ that are equivalent to ϕ'.

Note that in the presence of representations equivalent to ϕ' the projection on the RHS of the projection formula projects on all subrepresentations that are equivalent to ϕ' rather than just ϕ'. This need to be taken into account when performing character randomized benchmarking in general.

Pauli transfer matrix representation of quantum channels

Quantum channels [36, 38] are completely positive and trace-preserving (CPTP) linear maps $\mathcal{E} : \mathcal{M}_d \to \mathcal{M}_d$ where $\mathcal{M}_d$ is the Hilbert space of $d \times d$ Hermitian matrices. We will denote quantum channels by calligraphic font throughout. The canonical example of a quantum channel is conjugation by a unitary U, which we denote by the corresponding calligraphic letter, i.e., $\mathcal{U}(\rho) = U\rho U^\dagger$ for all density matrices ρ. We will denote the noisy implementation of a channel by an overset tilde, e.g., $\widetilde{\mathcal{G}}$ denotes a noisy implementation some ideal quantum channel channel $\mathcal{G}$.

It is often useful to think of quantum channels as matrices acting on vectors [35, 36, 39]. In order to do this we must choose a basis of the vector space $\mathcal{M}_d$. A convenient basis (when $d = 2^q$) will be the basis of normalized Pauli matrices $\{\sigma_0\} \cup \boldsymbol{\sigma}_q$ where $\sigma_0 := \mathbb{1}_d/\sqrt{d}$ with $d = 2^q$ is the normalized identity matrix and

$$\boldsymbol{\sigma}_q := \left\{ \frac{1}{\sqrt{d}} \{\mathbb{1}_2, X, Y, Z\}^{\otimes q} \right\} \backslash \{\sigma_0\}, \tag{22}$$

where X, Y, Z are the standard single qubit Pauli matrices. This set spans $\mathcal{M}_d$ and becomes an orthonormal basis when we equip $\mathcal{M}_d$ with the trace (or Hilbert-Schmidt) inner product defined as

$$\langle A, B\rangle := \mathrm{Tr}(AB^\dagger), \qquad \forall A, B \in \mathcal{M}_d. \tag{23}$$

For any element A of $\mathcal{M}_d$ we will denote its vectorization as $|A\rangle$. $|A\rangle$ is a column vector of length d^2 obtained by considering the set $\{\sigma_0\} \cup \boldsymbol{\sigma}_q$ as a basis for $\mathcal{M}_d$, that is

$$|A\rangle = \sum_{\sigma \in \{\sigma_0\} \cup \boldsymbol{\sigma}_q} \langle A, \sigma\rangle |\sigma\rangle. \tag{24}$$

The column vector vectorization $|\ \rangle$ of $\mathcal{M}_d$ has a natural dual row-vector vectorization which we denote by $\langle\ |$.

As any quantum channel $\mathcal{E}$ is a linear map from $\mathcal{M}_d$ to $\mathcal{M}_d$ we have

$$|\mathcal{E}(\rho)\rangle = \sum_{\sigma \in \{\sigma_0\} \cup \boldsymbol{\sigma}_q} |\mathcal{E}(\sigma)\rangle\langle\sigma|\rho\rangle, \tag{25}$$

so that we can represent $\mathcal{E}$ by the matrix

$$\mathcal{E} = \sum_{\sigma \in \{\sigma_0\} \cup \boldsymbol{\sigma}_q} |\mathcal{E}(\sigma)\rangle\langle\sigma|, \tag{26}$$

where we abuse notation by using the same symbol to refer to an abstract channel and its matrix representation. We will call this matrix the Pauli Transfer Matrix representation of the channel $\mathcal{E}$. The action of a channel $\mathcal{E}$ on a density matrix ρ now corresponds to the standard matrix action on the vector $|\rho\rangle$, hence for a density matrix ρ and a POVM element Q in $\mathcal{M}_d$ we have

$$\mathcal{E}|\rho\rangle = |\mathcal{E}(\rho)\rangle, \tag{27}$$

$$\mathrm{Tr}(Q\mathcal{E}(\rho)) = \langle Q|\mathcal{E}|\rho\rangle. \tag{28}$$

The Pauli transfer matrix representation has the nice properties (as can be easily checked) that the composition of quantum channels is equivalent to matrix multiplication of their Liouville matrices and that tensor products of channels correspond to tensor products of the corresponding Liouville matrices, that is, for all channels $\mathcal{E}_1$ and $\mathcal{E}_2$ and all $A \in \mathcal{M}_d$,

$$|\mathcal{E}_1 \circ \mathcal{E}_2(A)\rangle = \mathcal{E}_1\mathcal{E}_2|A\rangle, \tag{29}$$

$$|\mathcal{E}_1 \otimes \mathcal{E}_2(A^{\otimes 2})\rangle = \mathcal{E}_1 \otimes \mathcal{E}_2|A^{\otimes 2}\rangle. \tag{30}$$

Another nice property of the Pauli transfer matrix representation is that it is an actual representation (in the sense of eq. (9)) of any subgroup of the unitary group $\mathbb{U}(2^q)$. This essentially follows from eq. (29). For $U, V \in U(d)$ set $W = UV$. We then have for the Pauli transfer matrix representation

$$\mathcal{U}\mathcal{V}|X\rangle = |\mathcal{U}(\mathcal{V}(X))\rangle = |VUXU^\dagger V^\dagger\rangle = |VUX(VU)^\dagger\rangle = |WXW^\dagger\rangle = |\mathcal{W}(X)\rangle = \mathcal{W}|X\rangle, \qquad \forall X \in \mathcal{M}_d, \tag{31}$$

which is essentially the definition of a representation.

The last important property of the Pauli Transfer Matrix of a quantum channel is the fact that for a CPTP quantum channel $\mathcal{E}$ the average fidelity to the identity $F_{\mathrm{avg}}(\mathcal{E}, \mathcal{I})$, which is defined as

$$F_{\mathrm{avg}}(\mathcal{E}, \mathcal{I}) = \int d\psi\, \mathrm{Tr}(|\psi\rangle\langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)), \tag{32}$$

where $|\psi\rangle\langle\psi|$ is the regular density matrix of the pure state ψ and the integral is taken over the Haar measure on the set of pure states, is related to the trace (taken over superoperators) of the Pauli transfer matrix of $\mathcal{E}$. In particular we have

$$\frac{1}{|\boldsymbol{\sigma}_q|} \sum_{\sigma \in \boldsymbol{\sigma}_q} \langle\sigma|\mathcal{E}|\sigma\rangle = \frac{d-1}{d} F_{\mathrm{avg}}(\mathcal{E}, \mathcal{I}). \tag{33}$$

This means we can calculate the average fidelity to the identity of a quantum channel by calculating the trace over a subrepresentation (spanned by σ_q) of the Pauli transfer matrix representation. This observation which was first proven in [40], together with Schur's lemma, is at the core of the randomized benchmarking procedure. We will give a short proof of eq. (33) for completeness. We begin by proving it for the depolarizing channel. Let $\mathcal{D}_p$ be the *depolarizing* channel with depolarization probability p, that is

$$\mathcal{D}_p(X) = pX + (1-p)\operatorname{Tr}(X)\frac{\mathbb{1}}{d}. \tag{34}$$

Noting that $\operatorname{Tr}(|\psi\rangle\langle\psi|) = \operatorname{Tr}(|\psi\rangle\langle\psi|^2) = 1$ we see that

$$F_{\text{avg}}(\mathcal{D}_p, \mathcal{I}) = \int d\psi \operatorname{Tr}(|\psi\rangle\langle\psi|\mathcal{D}_d(|\psi\rangle\langle\psi|)) \tag{35}$$

$$= \int d\psi \left(p\operatorname{Tr}(|\psi\rangle\langle\psi|^2) + (1-p)\operatorname{Tr}(|\psi\rangle\langle\psi|/d)\operatorname{Tr}(|\psi\rangle\langle\psi|)\right) \tag{36}$$

$$= p + \frac{(1-p)}{d}, \tag{37}$$

which gives us the RHS of eq. (33). To compute the LHS we note that for $\sigma \in \{\sigma_0\} \cup \boldsymbol{\sigma}_q$ we have

$$\langle\sigma|\mathcal{D}_p|\sigma\rangle = p\operatorname{Tr}(\sigma^2) + (1-p)\operatorname{Tr}(\sigma/d) = p + \frac{(1-p)}{\sqrt{d}}. \tag{38}$$

Plugging this into the LHS of eq. (33) the equality becomes manifest. Now let us generalize to arbitrary quantum channels. Let $\mathcal{E}$ be an arbitrary channel. The first thing we notice is that F_{avg} is a covariant quantity, that is we have for all $U \in \mathbb{U}(d)$ that

$$F_{\text{avg}}(\mathcal{U}^\dagger\mathcal{E}\mathcal{U}, \mathcal{I}) = \int d\psi \operatorname{Tr}(|\psi\rangle\langle\psi|U^\dagger\mathcal{E}(U|\psi\rangle\langle\psi|U^\dagger)U) = \int \operatorname{Tr}(|\psi\rangle\langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)) = F_{\text{avg}}({}^\dagger\mathcal{E}, \mathcal{I}), \tag{39}$$

where we used the fact that the uniform measure over states is invariant under any unitary U and the cyclicity of the trace. One can similarly prove that

$$\frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{E}|\sigma\rangle = \frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{U}^\dagger\mathcal{E}\mathcal{U}|\sigma\rangle, \tag{40}$$

for all $U \in \mathbb{U}(d)$. A final statement we need is that, for all channels $\mathcal{E}$ the averaged channel $\int dU\mathcal{U}^\dagger\mathcal{E}\mathcal{U}$ is a depolarizing channel. This is proven in e.g. [3]. By linearity this means we have

$$\frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{E}|\sigma\rangle = \int dU\frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{U}^\dagger\mathcal{E}\mathcal{U}|\sigma\rangle = \int dU F_{\text{avg}}(\mathcal{U}\mathcal{E}\mathcal{U}, \mathcal{I}) = F_{\text{avg}}(\mathcal{E}, \mathcal{I}), \tag{41}$$

which proves the statement. We can generalize the above to give a powerful statement about randomized benchmarking procedures. Let $\mathbb{G}$ be a group and let $\mathcal{G} = \sum_\lambda \phi_\lambda(G)$ be the decomposition of the PTM representation into irreducible subrepresentations. For simplicity we will assume that there are no equivalent representations present in the above decomposition. Now consider for any quantum channel $\mathcal{E}$ the averaged quantum channel

$$\frac{1}{|\mathbb{G}|}\sum_{G\in\mathbb{G}}\mathcal{G}^\dagger\mathcal{E}\mathcal{G}. \tag{42}$$

Using lemma 1 we see that the above decomposes as

$$\frac{1}{|\mathbb{G}|}\sum_{G\in\mathbb{G}}\mathcal{G}^\dagger\mathcal{E}\mathcal{G} = \sum_\lambda \mathcal{P}_\lambda f_\lambda(\mathcal{E}), \tag{43}$$

where $\mathcal{P}_\lambda$ is the projector onto the representation ϕ_λ and f_λ is the associated quality parameter. Now noting that $\mathbb{G} \subset \mathbb{U}(d)$ we have that

$$\frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{E}|\sigma\rangle = \frac{1}{|\mathbb{G}|}\sum_{G\in\mathbb{G}}\langle\sigma|\mathcal{G}^\dagger\mathcal{E}\mathcal{G}|\sigma\rangle = \sum_\lambda f_\lambda\left(\frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{P}_\lambda|\sigma\rangle\right). \tag{44}$$

This, and eq. (33), implies that we can find the average fidelity of $\mathcal{E}$ by computing the weighted sum of the quality factors f_λ with the weights given by $\frac{1}{|\boldsymbol{\sigma}_q|}\sum_{\sigma\in\boldsymbol{\sigma}_q}\langle\sigma|\mathcal{P}_\lambda|\sigma\rangle = \frac{\operatorname{supp}(\phi_\lambda)}{|\boldsymbol{\sigma}_q|}$. This argument also holds in the presence of equivalent representations, see e.g. [21].

Representations of the Pauli group

Probably the most useful choice for the group $\hat{\mathbb{G}}$ is the multi-qubit Pauli group. This group is defined as $\mathbb{P} = \langle i\mathbb{1}, X, Z\rangle^{\otimes q}$. The reason this group is useful lies in the fact that the irreducible subrepresentations of the Pauli transfer matrix representations of $\mathbb{P}$ are all of dimension one and moreover that they are all inequivalent. In this section we will work out this fact more formally. Begin by considering the action of an element P, represented by $\mathcal{P}$, of the Pauli group $\mathbb{P}$ on an element σ of the basis $\{\sigma_0\} \cup \boldsymbol{\sigma}_q$.

$$\mathcal{P}|\sigma\rangle = |P\sigma P^\dagger\rangle = (-1)^{\langle P,\sigma\rangle}|P\rangle, \tag{45}$$

where $\langle A, B\rangle$ is a function that is zero whenever A, B commute and one if they anti-commute where A, B are (possibly normalized) elements of the Pauli group. Since (normalized) elements of the Pauli group always either commute or anti-commute this is a well defined function. Since the above is true for every element $P \in \mathbb{P}$ this implies that every element of $\{\sigma_0\} \cup \boldsymbol{\sigma}_q$ carries a one-dimensional, and hence irreducible, subrepresentation of the Pauli Transfer Matrix representations of the Pauli group. We will label these representations as ϕ_σ for all $\sigma \in \{\sigma_0\} \cup \boldsymbol{\sigma}_q$. Note that the character function χ_σ associated to the representation ϕ_σ is simply $\chi_\sigma(P) = (-1)^{\langle P,\sigma\rangle}$. This just follows from the definition of the character function, eq. (45) and the fact that $\langle\sigma|\sigma\rangle = 1$.

However, the representations ϕ_σ could a priori be equivalent. We now prove that this is not the case. To see this, recall the definition of equivalence from eq. (12). For one-dimensional vector spaces the only invertible linear transformation is multiplication by a non-zero constant. Hence if for two one-dimensional representations ϕ_σ and $\phi_{\sigma'}$ there exists a linear transformation such that $T \circ \phi_\sigma(P) = \phi_{\sigma'}(P) \circ T$ for all $P \in \mathbb{P}$ then $\phi_\sigma(P)\phi_{\sigma'}(P)$ for all $P \in \mathbb{P}$. This means, in order to prove that ϕ_σ and $\phi_{\sigma'}$ are inequivalent we simply need to find a $P \in \mathbb{P}$ such that $\phi_\sigma(P) \neq \phi_{\sigma'}(P)$. This comes down to finding, for each pair σ, σ' with $\sigma \neq \sigma'$, a $P \in \mathbb{P}$ such that $\mathcal{P}|\sigma\rangle = -\mathcal{P}|\sigma'\rangle$. This comes down to finding a Pauli element that commutes with σ and anti-commutes with σ'. It turns out this is always possible (and a little tedious to prove). For a proof of this fact see for instance [41, Lemma 7].

Note that for two Pauli matrices P, P' we can also efficiently (in the number of qubits in the system) decide whether they commute or anti-commute. This means that the character function $\chi_\sigma(P)$ can be efficiently computed on the fly for any σ and P. This is important because we must compute an instantiation of the character function for every random sample drawn during the character randomized benchmarking procedure. Note however that this can be done in post-processing so high speed (not just efficient) calculation of the character function is not a requirement for the success of the character randomized benchmarking procedure.

APPENDIX B: FINITE SAMPLING

In this section we elaborate on the statistical aspects of character randomized benchmarking. We will denote probability distributions by capital Greek letters (such as Λ) and their means by the letter μ subscripted by the corresponding distribution. The character randomized benchmarking protocol requires one to calculate the means of probability distributions. This is however impossible to do exactly using only a finite amount of samples drawn from his probability distribution. Instead one must rely on empirical estimates of these means. The reliability of these estimates is expressed by *confidence intervals*. Imagine being given a distribution with mean μ and an empirical estimate $\mu_N = \frac{1}{N}\sum_{x\in R_N} x$ where R_N is a set of N samples drawn independently from the distribution. Now a confidence interval (around μ_N) is a pair of real numbers (ϵ, δ) such that

$$\Pr(|\mu_N - \mu| \geq \epsilon) \leq 1 - \delta, \tag{46}$$

where the probability is taken with respect to the distribution being sampled from. Even though confidence intervals seem to require intimate knowledge of the distribution being sampled from they can in fact be constructed using only very limited knowledge of the distribution. In particular, if one knows that the distribution being sampled from is bounded, that is it only takes value inside an interval $[a, b]$ for $a, b \in \mathbb{R}$ then we can use Hoeffding's concentration inequality [34], given by

$$\Pr(|\mu_N - \mu| \geq \epsilon) \leq 1 - 2\exp\left(\frac{-N\epsilon^2}{(a-b)^2}\right). \tag{47}$$

Plugging in δ and inverting this equation we get a relation between the confidence interval (ϵ, δ) and the number of samples N from the distribution we need to construct this interval. We have

$$N \geq \frac{\log(2/\delta)(a-b)^2}{\epsilon^2}. \tag{48}$$

Note that this equation is completely generic, it can be used to empirically estimate the mean of any probability distribution, as long as this distribution is bounded.

With the above we can analyze the character randomized benchmarking protocol in the presence of finite sampling. The main question we aim to answer here is how many samples are required to accurately estimate the character average k_m^λ for fixed m and λ. There are 3 sources of randomness in the character randomized benchmarking protocol.

1. The first source of randomness comes from sampling sequences uniformly at random from the set $\mathbb{G}^{\times m}$
2. The second source of randomness comes from sampling an element from $\hat{\mathbb{G}}$ uniformly at random.
3. The last source of randomness is quantum mechanics itself. In general we can perform the following sequence of events
   (a) Prepare a system in a state ρ
   (b) Apply some quantum operation $\mathcal{E}$
   (c) Measure using some two-component POVM $\{Q, \mathbb{1} - Q\}$.

   At the end of this sequence we will get a single bit of information x which takes the value 0 (measure Q) or 1 (measure $\mathbb{1} - Q$). We can think of x as the being an instance of a random variable X which follows a Bernoulli distribution Λ_{Bern} with mean $\mu_{\Lambda_{\text{Bern}}} = \langle Q|\mathcal{E}|\rho\rangle$.

As mentioned in the main text, one of the key challenges of character randomized benchmarking lies in estimating the mean of the distribution induced by uniform random sampling from the group $\hat{\mathbb{G}}$ (the second source of randomness). Formally we have

$$k_m^\lambda(\vec{G}) = \frac{1}{|\hat{\mathbb{G}}|} \sum_{\hat{G} \in \hat{\mathbb{G}}} \chi_\lambda(\hat{G}) \operatorname{Tr}(P_\lambda) \langle Q|\widetilde{\mathcal{G}}_{\text{inv}} \widetilde{\vec{\mathcal{G}}} \hat{\mathcal{G}}|\rho\rangle \tag{49}$$

Note that this quantity mixes two of the above types of randomness as $k_m^\lambda(\vec{G})$ is an average of quantities $\langle Q|\widetilde{\mathcal{G}}_{\text{inv}} \widetilde{\vec{\mathcal{G}}} \hat{\mathcal{G}}|\rho\rangle$ which are themselves means of Bernoulli distributions.

The naive way of estimating $k_m^\lambda(\vec{G})$ would be to first estimate the means $\langle Q|\widetilde{\mathcal{G}}_{\text{inv}} \widetilde{\vec{\mathcal{G}}} \hat{\mathcal{G}}|\rho\rangle$ by performing the associated measurement procedure N times and using the concentration inequality given above to construct an (accurate) estimate of $\langle Q|\widetilde{\mathcal{G}}_{\text{inv}} \widetilde{\vec{\mathcal{G}}} \hat{\mathcal{G}}|\rho\rangle$. We can then multiply each estimate by $\chi_\lambda(\hat{G}) \operatorname{Tr}(P_\lambda)$ and average them to obtain an estimate for $k_m^\lambda(\vec{G})$.

However, to calculate $k_m^\lambda(\vec{G})$ we would have to perform this procedure for every $\hat{G} \in \hat{\mathbb{G}}$, which would require $|\hat{\mathbb{G}}|N$ samples in total. This is not a good approach when performing character randomized benchmarking on more than a few qubits. The reason for this is that typically the size of $\hat{\mathbb{G}}$ will grow exponentially with the number of qubits. For instance, if $\hat{\mathbb{G}}$ is the Pauli group we have $|\hat{\mathbb{G}}| = |\mathbb{P}| = 4^n$ for n qubits.

A second method, which will be more efficient when $|\hat{\mathbb{G}}|$ is very big, is to not try to estimate all means $\langle Q|\widetilde{\mathcal{G}}_{\text{inv}} \widetilde{\vec{\mathcal{G}}} \hat{\mathcal{G}}|\rho\rangle$ individually. Instead we will perform an empirical estimate of $k_m^\lambda(\vec{G})$ directly by the following procedure.

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

Every time we perform steps (1)-(3) we are are drawing a single sample from a certain probability distribution. This probability distribution is a *mixture distribution*. Mixture distributions are defined as linear combinations of probability distributions. Note that there there is a difference between a mixture of distributions and an linear combination of random variables [42]. Formally the mixture distribution induced by the procedure outlined above will be defined as

$$\Lambda_\lambda = \frac{1}{|\hat{\mathbb{G}}|}\sum_{\hat{G}\in\hat{\mathbb{G}}} \mathrm{Tr}(P_\lambda)\chi_\lambda(\hat{G})\Lambda_{\mathrm{Bern},\hat{G}} \tag{50}$$

where $\Lambda_{\mathrm{Bern},\hat{G}}$ is a Bernoulli distribution with mean $\mu_{\Lambda_{\mathrm{Bern},\hat{G}}} = \langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle$. The distribution Λ_λ will in general be rather complex (as it is the mixture of $|\hat{\mathbb{G}}|$ Bernoulli distributions). A useful feature of mixture distributions however, is that their mean is given by the weighted average the means of the mixing distributions with the weights precisely given by the weights in the mixture [42]. In particular that means we have for μ_{Λ_λ} that

$$\mu_{\Lambda_\lambda} = \frac{1}{|\hat{\mathbb{G}}|}\sum_{\hat{G}\in\hat{\mathbb{G}}} \mathrm{Tr}(P_\lambda)\chi_\lambda(\hat{G})\mu_{\Lambda_{\mathrm{Bern},\hat{G}}} \tag{51}$$

$$= \frac{1}{|\hat{\mathbb{G}}|}\sum_{\hat{G}\in\hat{\mathbb{G}}} \mathrm{Tr}(P_\lambda)\chi_\lambda(\hat{G})\langle Q|\widetilde{\mathcal{G}}_{\mathrm{inv}}\widetilde{\vec{\mathcal{G}}}\hat{\mathcal{G}}|\rho\rangle \tag{52}$$

$$= k^\lambda_m(\vec{G}). \tag{53}$$

Moreover the distribution Λ_λ is upper and lower bounded by $\pm\,\mathrm{Tr}(P_\lambda)\chi{*}_\lambda$ where $\chi^*_\lambda = \max_{\hat{G}}\chi_\lambda(\hat{G})$. This means that we can use the concentration inequality eq. (47) to bound the number of times we need to sample from Λ_λ (via the procedure above) in order to estimate $k^\lambda_m(\vec{G})$. Note that the number of samples that need to be taken will now not depend on $|\hat{\mathbb{G}}|$ at all.

As an illustration consider the follow example. Let $\hat{\mathbb{G}}$ be the Pauli group $\mathbb{P}$ on n qubits. This group is of size $|\mathbb{P}| = 4^n$. However, as discussed above, the subrepresentations of of the Pauli transfer matrix representation $\mathcal{P}$ are all of dimension one and are indexed by the normalized Pauli matrices $\sigma \in \{\sigma_0\}\cup\boldsymbol{\sigma}_q$. Let's perform character randomized benchmarking where $\lambda = \sigma$ for some normalized Pauli matrix σ. Since the representation ϕ_σ is one dimensional we have $\mathrm{Tr}(P_\sigma) = 1$. Moreover we have that the character $|\chi_\sigma(P)| = 1$ for all $P \in \mathbb{P}$. This means that the distribution Λ_σ is upper and lower bounded by ± 1. If we now want to estimate the mean $k^\lambda_m(\vec{G})$ for a particular sequence $\vec{G}$ we can perform the procedure above to sample from Λ_σ. Using the concentration inequality eq. (47) see that for a confidence interval of size $\epsilon = 0.02$ and confidence $\delta = 0.99$ around the mean $\mu_{\Lambda_\sigma} = k^\lambda_m(\vec{G})$ we need to draw

$$N \geq \frac{\log(2/0.99)(1-(-1))^2}{0.02^2} = 1769 \tag{54}$$

samples. Note that this number is both 'reasonable' and completely independent of the number of qubits n.

We make a final note about step (1) in the procedure for estimating $\langle Q|\mathcal{E}|\rho\rangle$, that is the preparation of the state ρ. It will often be the case that the optimal state for a character randomized benchmarking procedure, is not a pure state but rather represented by a density matrix of high rank. This introduces further experimental difficulties as an experimental setup usually only gives access to pure states (by design). We can overcome this difficulty by realizing that every density matrix ρ can be written as a probability distribution over pure states, that is

$$\rho = \sum_\psi p^\rho_\psi |\psi\rangle\langle\psi|, \qquad p^\rho_\psi \geq 0, \qquad \sum_\psi p^\rho_\psi = 1. \tag{55}$$

This means that $\langle E|\mathcal{E}|\rho\rangle$ is also the mean of a mixture distribution that takes values in the set $\{0,1\}$ (so the mixture is still a Bernoulli distribution).In particular it is a mixture of Bernoulli distributions with mean $\langle Q|\mathcal{E}|\psi\rangle$. This means that in the case of non-pure ρ we can update our sampling procedure to be

1. Fix a decomposition $\rho = \sum_\psi p^\rho_\psi|\psi\rangle\langle\psi|$
2. Sample ψ according to $\{p^\rho_\psi\}_\psi$
3. Sample $\hat{G} \in \hat{\mathbb{G}}$ uniformly at random

4. Prepare the state $\mathcal{G}_{\text{inv}}\mathcal{G}_m \cdots \mathcal{G}_1 \hat{\mathcal{G}}|\psi\rangle$ and measure it once obtaining a result $b(\hat{G}) \in \{0, 1\}$
5. Compute
   $x(\hat{G}) \;=\; \chi_{\hat{\lambda}'}(\hat{G}) \operatorname{Tr}(\mathcal{P}_{\hat{\lambda}'}) b(\hat{G}) \;\in\; \{0, \chi_{\hat{\lambda}'}(\hat{G}) \operatorname{Tr}(\mathcal{P}_{\hat{\lambda}'})\}$
6. Repeat sufficiently many times and compute the empirical average of $x(\hat{G})$.

This means we are now sampling from the mixture distribution

$$\Lambda_\lambda = \frac{1}{|\hat{\mathbb{G}}|} \sum_{\hat{G} \in \hat{\mathbb{G}}} \sum_{\psi} p_\psi^\rho \operatorname{Tr}(P_\lambda) \chi_\lambda(\hat{G}) \Lambda_{\text{Bern},\hat{G},\psi} \tag{56}$$

where $\Lambda_{\text{Bern},\hat{G},\psi}$ is now a Bernoulli distribution with mean $\langle Q|\mathcal{E}|\psi\rangle$. However the same reasoning as above holds and the number of samples (repetitions of the above procedure) required to obtain an estimate for the mean of Λ_λ still only depends on the interval on which Λ_λ is defined, yielding no increase in the number of samples needed even when the ideal input state ρ is very non-pure (has high rank).