Randomized benchmarking with gate-dependent noise

Joel J. Wallman

Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

December 18, 2017

We analyze randomized benchmarking for arbitrary gate-dependent noise and prove that the exact impact of gate-dependent noise can be described by a single perturbation term that decays exponentially with the sequence length. That is, the exact behavior of randomized benchmarking under general gate-dependent noise converges exponentially to a true exponential decay of exactly the same form as that predicted by previous analysis for gate-independent noise. Moreover, we show that the operational meaning of the decay parameter for gate-dependent noise is essentially unchanged, that is, we show that it quantifies the average fidelity of the noise between ideal gates. We numerically demonstrate that our analysis is valid for strongly gate-dependent noise models. We also show why alternative analyses do not provide a rigorous justification for the empirical success of randomized benchmarking with gate-dependent noise.

1 Introduction

The development of practical, large-scale devices that process quantum information relies upon techniques for efficiently characterizing the quality of control operations on the quantum level. Fully characterizing quantum processes [1, 2] requires resources that scale exponentially with the number of qubits despite improvements such as compressed sensing and direct fidelity estimation [3, 4, 5, 6, 7]. However, some figures of merit can be efficiently estimated without fully characterizing quantum processes. The canonical such figure is the fidelity of an experimental implementation $\tilde{\mathcal{G}}$ of an ideal unitary channel $\mathcal{G}(\rho) = G\rho G^\dagger$ averaged over input states distributed according to the Haar measure $d\psi$,

$$f(\tilde{\mathcal{G}}, \mathcal{G}) = \int d\psi \, \mathrm{Tr}[\mathcal{G}(\psi)\tilde{\mathcal{G}}(\psi)]. \tag{1}$$

At present, the only efficient partial characterizations are randomized benchmarking (RB) [8, 9, 10, 11, 12] and variants thereof [13, 14, 15, 16, 17, 18, 19]. The RB protocol of Ref. [12] provides an estimate of the average fidelity over a group of operations from how the fidelity of sequences of operations decays as more operations are applied. RB protocols scale efficiently with the number of qubits and are robust to state-preparation and measurement (SPAM) errors. Consequently, RB protocols have become an important baseline for the validation and verification of quantum operations [20, 21] and have recently been used to efficiently optimize over experimental control procedures [22].

However, rigorous analyses of RB protocols make several key assumptions, namely, that the noise is Markovian and time- and gate-independent, so that RB decay curves provide somewhat heuristic estimates of noise parameters (as do all noise characterization protocols). Numerical simulations of RB experiments have shown that the standard single-exponential decay curve holds (approximately) for a variety of physically relevant noise models with gate-dependent and non-Markovian noise, although the 'true' and estimated fidelity often differ by a factor of approximately two [23, 25]. However, it has recently been noted that gate-dependent noise fluctuations can dominate the gate-independent decay curve, resulting in a different decay parameter than expected when compared to a standard definition of the average noise [26].

In this paper, we prove that, under the assumption of Markovian noise, the full effect of gate-dependent fluctuations can be described by a single perturbation term that decays exponentially with the sequence length m. We also prove that the RB decay parameter is the fidelity of a suitably defined average noise process $\mathcal{E}$. Informally, our main result (theorem 4) is that for gate-dependent, but trace-perserving and Markovian noise, the average survival probability over all RB sequences of length m is

$$Ap(\mathcal{E})^m + B + \epsilon_m \tag{2}$$

for some constants A and B, where

$$\begin{aligned} p(\mathcal{E}) &= \frac{df(\mathcal{E},\mathcal{I}) - 1}{d-1} \\ |\epsilon_m| &\leq \delta_1 \delta_2^m, \end{aligned} \tag{3}$$

δ_1, δ_2 quantify the gate-dependence of the noise and δ_2 is small for good implementations of gates by theorem 3. We prove all theorems without restricting to trace-preserving noise, as we anticipate that near-term implementations of RB on encoded qubits [27] will involve significant loss from the encoded space due to imperfect error-correction procedures. We also demonstrate that the alternative analyses of Ref. [12, 28, 24, 26] do not rigorously justify fitting RB decays to a single exponential in section 7.

This paper is structured as follows. We introduce notation in section 2 and review the RB protocol of Ref. [12] in section 3. We then discuss an ambiguity in the decomposition of noise in section 4 and show how it can be exploited to cancel out the majority of gate-dependent fluctuations in section 5. We illustrate the accuracy of our new analysis numerically in section 6 and show why alternative analyses of RB for gate-dependent noise do not justify fitting RB experiments to single exponential decays in section 7. We conclude by discussing some implications of our results for implementing and interpreting RB.

2 Notation

We use the following notation throughout this paper.

- All operators except density operators are denoted by Roman font (e.g., A). The normalized identity matrix is $\hat{I}_d = I_d/\sqrt{d}$.

- Ideal channels are denoted by calligraphic font (e.g., $\mathcal{C}$), where the ideal channel $\mathcal{U}$ corresponding to a unitary operator U is $\mathcal{U}(\rho) = U\rho U^\dagger$.
- The unital component of a trace-preserving map is the linear map defined by $\mathcal{G}_{\rm u}(A) = \mathcal{G}(A - {\rm Tr}(A)/d)$.
- A noisy implementation of an ideal channel is denoted with an overset $\tilde{}$ (e.g., $\tilde{\mathcal{C}}$ denotes the noisy implementation of $\mathcal{C}$).
- Channel composition is denoted by (noncommutative) multiplication (i.e., $\mathcal{C}\mathcal{B}(A) = \mathcal{C}[\mathcal{B}(A)]$).
- Noncommutative products of subscripted objects are denoted by the shorthand

$$x_{b:a} = \begin{cases} x_b x_{b-1} \ldots x_{a+1} x_a & b \geq a, \\ 1 & \text{otherwise.} \end{cases} \tag{4}$$

- Groups and sets will be denoted by blackboard font, e.g., $\mathbb{G}$.
- The uniform average over a set $\mathbb{X}$ is denoted by $\mathbb{E}_{x\in\mathbb{X}}[f(x)] = |X|^{-1}\sum_{x\in\mathbb{X}} f(x)$.
- The trace and operator norms of a matrix are $\|M\|_1 = {\rm Tr}\sqrt{M^\dagger M}$ and $\|M\| = \max_v \|Mv\|/\|v\|$ respectively.
- The induced trace norm of a linear map $\mathcal{L}$ is

$$\|\mathcal{L}\|_{1\to 1} = \sup_{M:\|M\|_1=1} \|\mathcal{L}(M)\|_1, \tag{5}$$

  where the maximization is over positive-semidefinite matrices.

We use the following matrix representation of quantum channels (variously known as the Liouville, natural and Pauli transfer matrix representation) wherever a concrete representation is required. Let $\{e_1, \ldots, e_n\}$ be the canonical orthonormal unit basis of $\mathbb{C}^n$ and $\{B_1, \ldots, B_{d^2}\}$ be a trace-orthonormal basis of $\mathbb{C}^{d\times d}$, that is, ${\rm Tr}(B_j^\dagger B_k) = \delta_{j,k}$. Defining the linear map $|*\rangle\!\rangle : \mathbb{C}^{d\times d} \to \mathbb{C}^{d^2}$ by

$$|A\rangle\!\rangle = \sum_j {\rm Tr}(B_j^\dagger A) e_j \tag{6}$$

and the adjoint map $\langle\!\langle A| = |A\rangle\!\rangle^\dagger$, we have $\langle\!\langle A|B\rangle\!\rangle = {\rm Tr}(A^\dagger B)$. A channel $\mathcal{C}$ maps ρ to $\mathcal{C}(\rho)$, which can be represented by the matrix

$$\mathcal{C} = \sum_j |\mathcal{C}(B_j)\rangle\!\rangle\langle\!\langle B_j|, \tag{7}$$

where we abuse notation slightly by using the same notation for the abstract channel and its matrix representation. For the remainder of the paper, we assume the operator basis is Hermitian and that $B_1 = \hat{I}_d$, so that the matrix representation of all Hermiticity-preserving maps, including all completely positive and trace-preserving (CPTP) maps, are real-valued.

3 Randomized benchmarking protocol

The RB protocol for a group of operations $\mathbb{G}$ that is also a unitary two-design [11] (e.g., the n-qubit Clifford group) is as follows.

1. Choose a positive integer m.
2. Choose a sequence $\vec{G} = (G_1, \ldots, G_m) \in \mathbb{G}^m$ of m elements of $\mathbb{G}$ uniformly at random and set $G_{m+1} = (G_{m:1})^\dagger$.
3. Estimate the expectation value (also known as the survival probability) $Q_{\vec{G}} = \mathrm{Tr}[Q\tilde{\mathcal{G}}_{m+1:1}(\rho)]$ of an observable Q after preparing a d-level system in a state ρ and applying the sequence of $m+1$ operations $G_1, G_2, \ldots, G_{m+1}$.
4. Repeat steps 2–3 to estimate the average over random sequences $\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(Q_{\vec{G}})$ to a desired precision.
5. Repeat steps 1–4 for different values of m and fit to the decay model
$$\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(Q_{\vec{G}}) = Ap^m + B \tag{8}$$
to estimate p [12].

Despite the ubiquitous usage of RB, best practice choices for the sequence lengths and the number of sequences at each length are still unknown, though see Refs. [23, 30, 31] for some guidance.

A unitary two-design is any set of unitary channels $\mathbb{G}$ such that uniformly sampling $\mathbb{G}$ reproduces the first and second moments of the Haar measure. The canonical example of a unitary two-design is the n-qubit Clifford group. The unitary two-design condition can also be stated in terms of how a general channel is 'twirled' by the group, where the twirl of a channel $\mathcal{C}$ over a group $\mathbb{G}$ is $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^\dagger\mathcal{C}\mathcal{G})$. A unitary two-design is any group $\mathbb{G}$ such that any CPTP channel is 'twirled' into a depolarizing channel [11, 32]

$$\mathcal{D}_p(\rho) = p\rho + \frac{1-p}{d}I_d. \tag{9}$$

Unitary two-designs also twirl a general linear map $\mathcal{C}$ into a depolarizing map composed with a channel that uniformly decreases the trace, that is, into a channel of the form

$$\mathcal{D}_{p,t}(\rho) = \frac{t}{d}I_d + p(\rho - \frac{1}{d}I_d), \tag{10}$$

where the values of p and t can be computed from the following lemma [30]. Note that p and t are functions from linear maps to the real numbers, however, we will often omit the argument when it is clear from the context.

Lemma 1. *For any unitary two-design $\mathbb{G}$ and channel $\mathcal{C}$,*

$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^\dagger\mathcal{C}\mathcal{G}) = \mathcal{D}_{p,t}, \tag{11}$$

where

$$\begin{aligned} t(\mathcal{C}) &= \frac{\mathrm{Tr}[\mathcal{C}(I_d)]}{d} \\ p(\mathcal{C}) &= \frac{\mathrm{Tr}(\mathcal{C}) - t}{d^2-1} = \frac{df(\mathcal{C}) - t}{d-1}. \end{aligned} \tag{12}$$

In theorem 4, we prove that eq. (8) holds, up to a small and exponentially decaying perturbation, for gate-dependent noise that is time-independent and Markovian, generalizing the original analysis to more realistic noise.

4 Representing gate-dependent noise

Let $\tilde{\mathcal{G}}$ be an experimental implementation of an ideal unitary channel $\mathcal{G}(\rho) = G\rho G^\dagger$ with Markovian noise. It is convenient to decompose $\tilde{\mathcal{G}}$ into an ideal gate and a fictitious noise process, for example, $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{E}$. However, we can also write $\tilde{\mathcal{G}} = \mathcal{L}_G\mathcal{G}\mathcal{R}_G$ for any suitable $\mathcal{L}_G$ and $\mathcal{R}_G$. We now identify an approximate decomposition that will greatly simplify the analysis of RB with gate-dependent noise.

Theorem 2. *Let $\mathbb{G}$ be a unitary two-design, $\{\tilde{\mathcal{G}} : G \in \mathbb{G}\}$ be a corresponding set of Hermiticity-preserving maps, and p and t be the largest eigenvalues of $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}} \otimes \tilde{\mathcal{G}})$ and $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}})$ in absolute value respectively, where $\mathcal{G}_{\mathrm{u}}(A) = \mathcal{G}(A - \mathrm{Tr}(A)/d)$. There exist linear maps $\mathcal{L}$ and $\mathcal{R}$ such that*

$$\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{L}\mathcal{D}_{p,t} \tag{13a}$$

$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^\dagger\mathcal{R}\tilde{\mathcal{G}}) = \mathcal{D}_{p,t}\mathcal{R} \tag{13b}$$

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

Remark. When the noise is independent of the gate, that is, $\tilde{\mathcal{G}} = \mathcal{L}\mathcal{G}\mathcal{R}$ for all $G \in \mathbb{G}$ for some $\mathcal{L}$ and $\mathcal{R}$, then $\mathcal{L}$ and $\mathcal{R}$ are solutions to eq. (13), as can be seen by substituting $\tilde{\mathcal{G}} = \mathcal{L}\mathcal{G}\mathcal{R}$ into eq. (13) and using lemma 1.

For gate-dependent noise such that $\mathcal{R}$ is invertible, we can set $\tilde{\mathcal{G}} = \mathcal{L}_G\mathcal{G}\mathcal{R}$, so that the noise between two ideal gates $\mathcal{G}$ and $\mathcal{H}$ is $\mathcal{R}\mathcal{L}_G$. Substituting $\tilde{\mathcal{G}} = \mathcal{L}_G\mathcal{G}\mathcal{R}$ into eq. (13b) and multiplying by $\mathcal{R}^{-1}$ gives

$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^\dagger\mathcal{R}\mathcal{L}_G\mathcal{G}) = \mathcal{D}_{p,t}. \tag{14}$$

As the Haar measure is unitarily invariant and p and t are convex functions,

$$\begin{aligned} p &= p(\mathbb{E}_{G\in\mathbb{G}}[\mathcal{G}^\dagger\mathcal{R}\mathcal{L}_G\mathcal{G}]) = \mathbb{E}_{G\in\mathbb{G}}(p[\mathcal{R}\mathcal{L}_G]) \\ t &= t(\mathbb{E}_{G\in\mathbb{G}}[\mathcal{G}^\dagger\mathcal{R}\mathcal{L}_G\mathcal{G}]) = \mathbb{E}_{G\in\mathbb{G}}(t[\mathcal{R}\mathcal{L}_G]) \end{aligned} \tag{15}$$

give the fidelity and loss of trace of the average noise between ideal gates to the identity via lemma 1. If $\mathcal{R}$ is not invertible, we can introduce an arbitrarily small perturbation into the $\tilde{\mathcal{G}}$ to make $\mathcal{R}$ invertible and so eq. (15) will hold to an arbitrary precision.

When $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{D}_{p,t}$, $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}} \otimes \tilde{\mathcal{G}})$ and $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}})$ each have only one nonzero eigenvalue, namely, p and t respectively [15, 16]. By the Bauer-Fike theorem [33], all the eigenvalues η of a matrix $M + \delta M$ satisfy

$$\min_{\lambda\in\mathrm{spec}(M)} |\eta - \lambda| \leq \|\delta M\|, \tag{16}$$

where $\mathrm{spec}(M)$ is the spectrum of M. As the operator norm is submultiplicative, obeys the triangle inequality, and $\|A \otimes B\| = \|A\|\|B\|$, the largest eigenvalues of $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}} \otimes \tilde{\mathcal{G}})$ and $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}})$ will differ from p' and t' by at most $\mathbb{E}_{G\in\mathbb{G}}\|\tilde{\mathcal{G}} - \mathcal{G}\mathcal{D}_{p',t'}\|$ for any scalars p' and t'. Therefore the largest eigenvalues in absolute value will generally be unique and hence must be real as they are the eigenvalues of real matrices.

Proof. Equations (13a) and (13b) are essentially a pair of eigenvalue equations, except that $\mathbb{G}$ acts irreducibly on $|\hat{I}_d\rangle\rangle$ and its orthogonal complement. To utilize this structure, let $\mathcal{L} = |L\rangle\rangle\langle\langle\hat{I}_d| + \mathcal{L}'$ and $\mathcal{R} = |\hat{I}_d\rangle\rangle\langle\langle R| + \mathcal{R}'$ with $\mathcal{L}'|\hat{I}_d\rangle\rangle = 0$ and $\langle\langle\hat{I}_d|\mathcal{R}' = 0$ without loss of generality. With these definitions and as $\mathcal{G}|\hat{I}_d\rangle\rangle = |\hat{I}_d\rangle\rangle$ for any unitary (or unital) channel, we can separate each equation in eqs. (13a) and (13b) into two independent equations

$$\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}})|L\rangle\rangle = t|L\rangle\rangle \tag{17a}$$
$$\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}}\mathcal{L}'\mathcal{G}_{\mathrm{u}}^{\dagger}) = p\mathcal{L}' \tag{17b}$$
$$\langle\langle R|\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}}) = t\langle\langle R| \tag{17c}$$
$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}}^{\dagger}\mathcal{R}'\tilde{\mathcal{G}}) = p\mathcal{R}', \tag{17d}$$

where restricting to $\mathcal{G}_{\mathrm{u}}$ enforces the orthogonality constraints on $\mathcal{L}'|\hat{I}_d\rangle\rangle = 0$ and $\langle\langle\hat{I}_d|\mathcal{R}' = 0$. Equations (17a) and (17c) are left- and right-eigenvector problems of a real matrix and so t can be set to any eigenvalue of $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}})$ and L and R to the corresponding non-trivial eigenvectors (recall that the set of left- and right-eigenvalues of a real matrix are identical). We choose the largest eigenvalue as it will result in the smallest $\Delta_G = \tilde{\mathcal{G}} - \mathcal{L}\mathcal{G}\mathcal{R}$.

To turn eqs. (17b) and (17d) into eigenvalue equations, we can use the vectorization map that maps a matrix to a vector by stacking the columns vertically, that is,

$$\mathrm{vec}(\sum_j v_j e_j^{\dagger}) = \sum_j e_j \otimes v_j. \tag{18}$$

The vectorization map satisfies the identity

$$\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B), \tag{19}$$

Applying eq. (19) to eqs. (17b) and (17d) and using $\mathcal{G}^T = \mathcal{G}^{\dagger}$ (as the matrix basis is Hermitian) gives

$$\begin{aligned}\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}} \otimes \tilde{\mathcal{G}})\,\mathrm{vec}(\mathcal{L}') &= p\,\mathrm{vec}(\mathcal{L}')\\ \mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}} \otimes \mathcal{G}_{\mathrm{u}})^T\,\mathrm{vec}(\mathcal{R}') &= p\,\mathrm{vec}(\mathcal{R}'),\end{aligned} \tag{20}$$

and so p can be set to any eigenvalue of $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}} \otimes \tilde{\mathcal{G}})$ and $\mathcal{L}'$ and $\mathcal{R}'$ to the corresponding non-trivial eigenvectors (or equivalently, to any eigenvalue of $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}} \otimes \mathcal{G}_{\mathrm{u}})$ as the two are related by a unitary transformation and so have the same spectrum).

The maps $\mathcal{L}$ and $\mathcal{R}$ identified are solutions to eigenvector equations and so are only determined up to a normalization constant. As $\mathcal{D}_{q,u}$ commutes with $\mathcal{G}$ for all scalars q and v and all $\mathcal{G} \in \mathbb{G}$, we can set $\mathcal{L} \to \mathcal{L}D_{q,u}$ and $\mathcal{R} \to \mathcal{D}_{r,v}\mathcal{R}$ to satisfy eq. (13c). □

While applying a gauge transformation $\tilde{\mathcal{G}} \to \mathcal{S}^{-1}\tilde{\mathcal{G}}\mathcal{S}$ changes $\mathcal{L}$ and $\mathcal{R}$ to $\mathcal{L} \to \mathcal{S}\mathcal{L}$ and $\mathcal{R} \to \mathcal{R}\mathcal{S}^{-1}$ respectively, the noise between ideal gates, namely, $\mathcal{R}\mathcal{L}$, is gauge invariant and so the present analysis is gauge-invariant unlike the original analysis of RB as shown in Ref. [26].

We now prove that the solutions $\mathcal{L}$ and $\mathcal{R}$ to eq. (13) can also be chosen to ensure that $\Delta_G = \tilde{\mathcal{G}} - \mathcal{L}\mathcal{G}\mathcal{R}$ is small when $\tilde{\mathcal{G}} \approx \mathcal{G}\mathcal{D}_{p,t}$. To prove this, it is convenient to transform to the gauge $\tilde{\mathcal{G}}^{\mathcal{I}}$ such that $\mathcal{L} = \mathcal{I}$ (i.e., set $\mathcal{S} = \mathcal{L}^{-1}$). In this gauge, eq. (13c) becomes

$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}\mathcal{R}\mathcal{G}^{\dagger}) = \mathcal{D}_{p,t}. \tag{21}$$

In particular, Δ_G is on the order of $\|\tilde{\mathcal{G}} - \mathcal{G}\mathcal{D}_{p,t}\| \in O(\sqrt{1-p})$.

Theorem 3. *Let $\mathbb{G}$ be a unitary two-design, $\{\tilde{\mathcal{G}} : G \in \mathbb{G}\}$ be a corresponding set of Hermiticity-preserving maps, and p and t be the largest eigenvalues of $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}_{\mathrm{u}} \otimes \tilde{\mathcal{G}})$ and $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}})$ in absolute value respectively, where $\mathcal{G}_{\mathrm{u}}(A) = \mathcal{G}(A - \mathrm{Tr}(A)/d)$. There exists a linear map $\mathcal{R}$ satisfying eq. (13) with $\mathcal{L} = \mathcal{I}$ such that*

$$\|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{R}\| \leq \|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{D}_{p,t}\| + \frac{\|\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\tilde{\mathcal{G}}^{\mathcal{I}}) - \mathcal{D}_{p,t}\|}{1 - \mathbb{E}_{G\in\mathbb{G}}(\|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{D}_{p,t}\|)\|\mathcal{D}_{1/p,1/t}\|}. \tag{22}$$

Proof. By the triangle inequality, submultiplicativity, and unitary invariance of the operator norm,

$$\begin{aligned}\|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{R}\| &\leq \|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{D}_{p,t}\| + \|\mathcal{G}(\mathcal{D}_{p,t} - \mathcal{R})\| \\ &\leq \|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{D}_{p,t}\| + \|\mathcal{D}_{p,t} - \mathcal{R}\|\end{aligned} \tag{23}$$

for all $G \in \mathbb{G}$.

Equation (23) holds for any $\mathcal{R}$. Now let $\mathcal{R}$ be a solution to eq. (13) for $\tilde{\mathcal{G}}^{\mathcal{I}}$ and expand $\mathcal{R} = \mathcal{D}_{p,t} + \mathcal{R}_1$ and $\tilde{\mathcal{G}}^{\mathcal{I}} = \mathcal{G}\mathcal{D}_{p,t} + \tilde{\mathcal{G}}_1$. Substituting these expansions into eq. (13b) and using eq. (21) gives

$$\begin{aligned}\mathcal{D}_{p,t}^2 + \mathcal{D}_{p,t}\mathcal{R}_1 &= \mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\mathcal{R}\mathcal{G}\mathcal{D}_{p,t}) + \mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\mathcal{R}\tilde{\mathcal{G}}_1) \\ &= \mathcal{D}_{p,t}^2 + \mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\mathcal{R}\tilde{\mathcal{G}}_1)\end{aligned} \tag{24}$$

Canceling the common term in eq. (24), multiplying both sides by $\mathcal{D}_{1/p,1/t}$ from the left, taking the operator norm and using the triangle inequality and submultiplicativity of the operator norm gives

$$\begin{aligned}\|\mathcal{R}_1\| &\leq \|\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\tilde{\mathcal{G}}_1)\| + \|\mathbb{E}_{G\in\mathbb{G}}(\mathcal{D}_{1/p,1/t}\mathcal{G}^{\dagger}\mathcal{R}_1\tilde{\mathcal{G}}_1)\| \\ &\leq \|\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\tilde{\mathcal{G}}_1)\| + \mathbb{E}_{G\in\mathbb{G}}(\|\tilde{\mathcal{G}}_1\|\|\mathcal{R}_1\|\|\mathcal{D}_{1/p,1/t}\mathcal{G}^{\dagger}\|).\end{aligned} \tag{25}$$

Rearranging and using the unitary invariance of the operator norm gives

$$\begin{aligned}\|\mathcal{R}_1\| &= \|\mathcal{D}_{p,t} - \mathcal{R}\| \\ &\leq \frac{\|\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\tilde{\mathcal{G}}_1)\|}{1 - \mathbb{E}_{G\in\mathbb{G}}(\|\tilde{\mathcal{G}}_1\|)\|\mathcal{D}_{1/p,1/t}\|} \\ &= \frac{\|\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^{\dagger}\tilde{\mathcal{G}}^{\mathcal{I}}) - \mathcal{D}_{p,t}\|}{1 - \mathbb{E}_{G\in\mathbb{G}}(\|\tilde{\mathcal{G}}^{\mathcal{I}} - \mathcal{G}\mathcal{D}_{p,t}\|)\|\mathcal{D}_{1/p,1/t}\|}.\end{aligned} \tag{26}$$

□

5 Analyzing RB with arbitrarily gate-dependent noise

Experimental implementations of RB almost always produce relatively clean exponential decays. We now explain this empirical success for general gate-dependent noise by proving that the average survival probability over RB sequences of length m is equal to the model of Ref. [28] for gate-independent noise with a single perturbation term for a suitably defined average noise model. Moreover, the decay constants correspond to the average loss of trace and fidelity of the noise between ideal gates via eq. (15). The perturbation term will generally be negligible for noise that is close to the identity by theorem 3 and for $m \geq 3$ (for numerical values for a single qubit, see fig. 1). Even under this condition, more significant (and potentially oscillating) perturbations are possible for $m \approx 1$, potentially explaining experimentally observed fluctuations as seen in, e.g., [20, Fig. 3b]. While we have not investigated how the perturbation terms scale with dimension, note that we only need $\delta_2 \lessapprox 1/2$ in order for the decay to be essentially a single exponential for $m \geq 10$, etc, as δ_2 is exponentially suppressed.

Theorem 4. *For any unitary two-design $\mathbb{G}$ and set of Hermiticitiy preserving linear maps $\{\tilde{\mathcal{G}} : G \in \mathbb{G}\}$, the average survival probability over all randomized benchmarking sequences of length m is*

$$\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(Q_{\vec{G}}) = Ap^m + Bt^m + \epsilon_m \tag{27}$$

for some constants A and B, where p and t give the fidelity and loss of trace of the average noise between ideal gates via eq. (15). Moreover, the perturbation term ϵ_m satisfies

$$|\epsilon_m| \leq \delta_1 \delta_2^m \tag{28}$$

for some δ_1 and δ_2 that quantify the amount of gate dependence.

Remark. Equation (27) reduces to the standard single-exponential model plus an exponentially decreasing perturbation when $\tilde{\mathcal{G}}$ is a CPTP map for all G as then $\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}}\otimes\mathcal{G})$ and $\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}\otimes\tilde{\mathcal{G}})$ are both CPTP maps and so all eigenvalues are in the unit disc and the eigenvalue with eigenvector close to $|\hat{I}_d\rangle\rangle$ is 1 [34].

The constants A and B and the perturbation term are

$$\begin{aligned} A &= \langle\langle Q|\mathcal{L}|\mathcal{R}(\rho) - I_d/d\rangle\rangle \\ B &= \langle\langle Q|\mathcal{L}|I_d/d\rangle\rangle \\ \epsilon_m &= \langle\langle Q|\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\Delta_{m+1:1})|\rho\rangle\rangle. \end{aligned} \tag{29}$$

with $\Delta_j = \Delta_{G_j} = \tilde{\mathcal{G}}_j - \mathcal{L}\mathcal{G}_j\mathcal{R}$ and where $\mathcal{L}$ and $\mathcal{R}$ are solutions to eq. (13). The quantities bounding the perturbation term are

$$\begin{aligned} \delta_1 &= \max_G \|\Delta_G\|_{1\to1}\|\rho\|_1\|Q\|_1 \\ \delta_2 &= \mathbb{E}_{G\in\mathbb{G}}(\|\Delta_G\|). \end{aligned} \tag{30}$$

which are in turn bounded in theorem 3 in the gauge where $\mathcal{L} = \mathcal{I}$. Note that $\|\rho\|_1, \|Q\|_1 \leq 1$ in a standard gauge, however, the gauge transformation required to set $\mathcal{L} = \mathcal{I}$ may change these values and introduce negative eigenvalues to ρ so that the maximization for the induced norm in δ_1 is not restricted to positive semi-definite inputs.

Proof. The average map over all randomized benchmarking sequences $\vec{G}$ of length m is

$$\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:1}). \tag{31}$$

Let $\mathcal{L}$ and $\mathcal{R}$ satisfy

$$\mathbb{E}_{G\in\mathbb{G}}(\tilde{\mathcal{G}}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{L}\mathcal{D}_{p,t} \tag{32a}$$

$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}^\dagger\mathcal{R}\tilde{\mathcal{G}}) = \mathcal{D}_{p,t}\mathcal{R} \tag{32b}$$

$$\mathbb{E}_{G\in\mathbb{G}}(\mathcal{G}\mathcal{R}\mathcal{L}\mathcal{G}^\dagger) = \mathcal{D}_{p,t} \tag{32c}$$

and define

$$\Delta_j = \Delta_{G_j} = \tilde{\mathcal{G}}_j - \mathcal{L}\mathcal{G}_j\mathcal{R}. \tag{33}$$

Then for any integer $1 \le j \le m+1$ we can write

$$\begin{aligned}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:j}\Delta_{j-1:1}) &= \mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:j+1}\mathcal{L}\mathcal{G}_j\mathcal{R}\Delta_{j-1:1}) + \mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:j+1}\Delta_j\Delta_{j-1:1}) \\ &= \mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:j+1}\mathcal{L}\mathcal{G}_j\mathcal{R}\Delta_{j-1:1}) + \mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:j+1}\Delta_{j:1}),\end{aligned} \tag{34}$$

where $\Delta_{0:1} = I_d = \tilde{\mathcal{G}}_{m+1:m+2}$ by convention. Moreover, as $G_{m+1:1} = I_d$ and any set of m elements of $\vec{G}$ are statistically independent and uniformly distributed, we can apply eq. (32a) recursively and use the fact that $\mathcal{D}_{p,t}$ commutes with unital channels to obtain

$$\begin{aligned}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:2}\mathcal{L}\mathcal{G}_1\mathcal{R}) &= \mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:2}\mathcal{L}[\mathcal{G}_{m+1:2}]^\dagger\mathcal{R}) \\ &= \mathcal{L}\mathcal{D}_{p,t}^m\mathcal{R}\end{aligned} \tag{35}$$

for $j = 1$. Similarly, we can use lemma 1 and eq. (32) for $j \ge 2$ to obtain

$$\begin{aligned}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}[\tilde{\mathcal{G}}_{m+1:j+1}\mathcal{L}\mathcal{G}_j\mathcal{R}\Delta_{j-1:1}] &= \mathbb{E}_{\vec{G}\in\mathbb{G}^m}[\tilde{\mathcal{G}}_{m+1:j+1}\mathcal{L}(\mathcal{G}_{m+1:j+1})^\dagger(\mathcal{G}_{j-2:1})^\dagger\mathcal{G}_{j-1}^\dagger\mathcal{R}\Delta_{j-1}\Delta_{j-2:1}] \\ &= \mathcal{L}\mathcal{D}_{p,t}^{m-j}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}[(\mathcal{G}_{j-2:1})^\dagger\mathcal{G}_{j-1}^\dagger\mathcal{R}(\tilde{\mathcal{G}}_{j-1} - \mathcal{L}\mathcal{G}_{j-1}\mathcal{R})\Delta_{j-2:1}] \\ &= \mathcal{L}\mathcal{D}_{p,t}^{m-j}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}[(\mathcal{G}_{j-2:1})^\dagger\mathbb{E}_{G_{j-1}}(\mathcal{G}_{j-1}^\dagger\mathcal{R}\tilde{\mathcal{G}}_{j-1} - \mathcal{D}_{p,t}\mathcal{R})\Delta_{j-2:1}] \\ &= 0.\end{aligned} \tag{36}$$

Therefore we can apply eqs. (34) and (35) and then recursively apply eqs. (34) and (36) to obtain

$$\begin{aligned}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}\tilde{\mathcal{G}}_{m+1:1} &= \mathcal{L}\mathcal{D}_{p,t}^m\mathcal{R} + \mathbb{E}_{\vec{G}\in\mathbb{G}^m}[\tilde{\mathcal{G}}_{m+1:2}\Delta_1] \\ &= \mathcal{L}\mathcal{D}_{p,t}^m\mathcal{R} + \mathbb{E}_{\vec{G}\in\mathbb{G}^m}[\Delta_{m+1:1}].\end{aligned} \tag{37}$$

The average survival probability over all random sequences is then

$$\begin{aligned}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}Q_{\vec{G}} &= \langle\!\langle Q|\mathcal{L}\mathcal{D}_{p,t}^m\mathcal{R}|\rho\rangle\!\rangle + \epsilon_m \\ &= t^m\langle\!\langle Q|\mathcal{L}|I_d/d\rangle\!\rangle + p^m\langle\!\langle Q|\mathcal{L}|\mathcal{R}(\rho) - I_d/d\rangle\!\rangle + \epsilon_m\end{aligned} \tag{38}$$

where

$$\epsilon_m = \langle\!\langle Q|\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\Delta_{m+1:1})|\rho\rangle\!\rangle. \tag{39}$$

We now bound the perturbation term. By the triangle inequality and submultiplicativity,

$$\begin{aligned}\|\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:1}) - \mathcal{L}\mathcal{D}_{p,t}^m\mathcal{R}\|_{1\to1} &= \|\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(\Delta_{m+1:1})\|_{1\to1} \\ &\le \mathbb{E}_{\vec{G}\in\mathbb{G}^m}\|\Delta_{m+1:1}\|_{1\to1} \\ &\le \mathbb{E}_{\vec{G}\in\mathbb{G}^m}\|\Delta_1\|_{1\to1}\prod_{j=m+1}^{2}\|\Delta_j\| \\ &\le \max_G\|\Delta_G\|_{1\to1}\mathbb{E}_{\vec{G}\in\mathbb{G}^m}\prod_{j=m+1}^{2}\|\Delta_j\| \\ &\le \max_G\|\Delta_G\|_{1\to1}[\mathbb{E}_H\|\Delta_H\|]^m.\end{aligned} \tag{40}$$

□

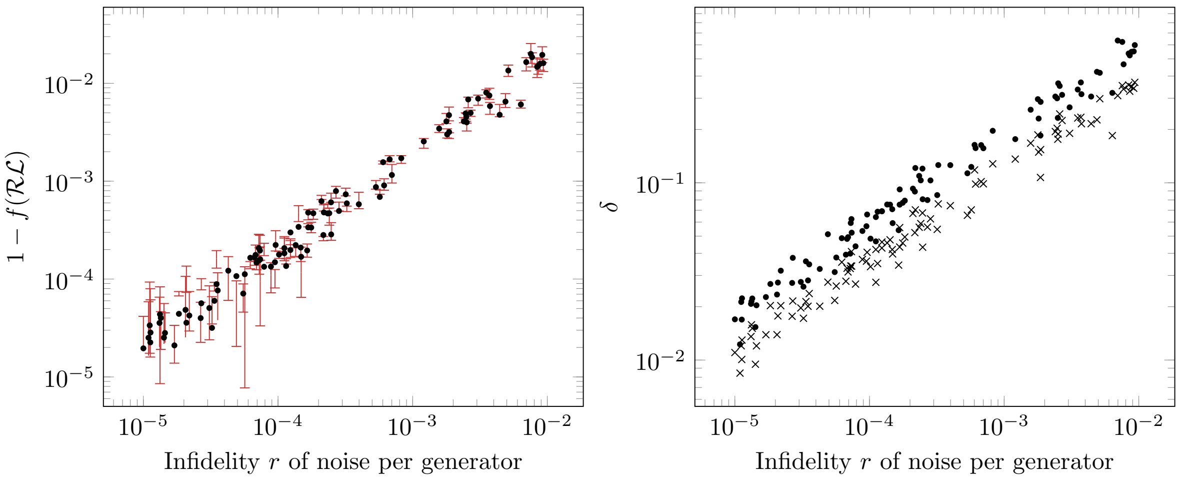

Figure 1: Left: upper- and lower- edges of 90% confidence intervals for the decay parameter from the numerical simulations of RB experiments described in section 6 fitted neglecting the perturbation term $\delta_1\delta_2^m$ (red intervals) and the corresponding values of p from theorem 4 (black dots). Right: numerical values of δ_1 (circles) and δ_2 (crosses) from theorem 4. For all simulated experiments, the perturbation term scales as $\delta_1\delta_2^m \in O(r^{-m/2})$ and so is negligible for $m \geq 3$.

6 Numerics

In fig. 1 we illustrate the reliability of our analysis by comparing the estimates obtained by fitting the simulated RB experiments described below under independent random unitary noise on a set of generating gates to the decay parameters predicted by theorem 4 and by plotting the size of the perturbation terms δ_1 and δ_2 calculated in the gauge where $\mathcal{L} = \mathcal{I}$. Mathematica code is available at https://github.com/jjwallman/numerics.

For simplicity, we simulate RB experiments using the (projective) group

$$\mathbb{G} = \{T^t P : t \in \mathbb{Z}_3, P \in \{I, X, Y, Z\}\} \tag{41}$$

where I, X, Y, Z are the standard single-qubit Pauli matrices and

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}, \tag{42}$$

which is a subgroup of the single-qubit Clifford group that is also a unitary two-design. For our simulations, the noisy implementation of $G = T^t P \in \mathbb{G}$ is $\tilde{\mathcal{G}} = \mathcal{G}'$ where

$$G' = UT^t VP \tag{43}$$

where U and V are independently sampled from the set $\mathbb{U}_r \subset \mathrm{SU}(d)$ satisfying $1 - f(\mathcal{U}, \mathcal{I}) = r$, that is, the set of unitary channels of fixed infidelity r. We then calculate $p(\mathcal{RL})$ from theorem 2 and δ_1, δ_2 from eq. (30). We set the state to $|0\rangle$ and the measurement to be a computational basis measurement and do not include SPAM errors for simplicity.

The set $\mathbb{U}_r$ can be sampled by sampling $V \in \mathrm{SU}(d)$ uniformly from the Haar measure and setting

$$U = V \exp(-i\theta Z) V^\dagger = \cos\theta I - i \sin\theta V Z V^\dagger, \tag{44}$$

which gives the unique unitarily invariant distribution over $\mathbb{U}_r$ for some r as a function of θ. In order for the infidelity of the resulting gates to be r, we require [32]

$$r = \frac{4 - |\mathrm{Tr}\, U|^2}{6} = \frac{2 - \cos^2\theta}{3}, \tag{45}$$

or $\theta = \arcsin(\sqrt{3r/2})$.

A single RB experiment for a fixed value of m then consisted of the following.

1. Randomly choose $\vec{a} \in \{0, 1, 2\}^m$ and $\vec{b} \in \{1, 2, 3, 4\}^m$.
2. Set $G_j = T^{a_j} P_{b_j} P_{b_{j-1}} T^{-a_{j-1}} = T^{a_j - a_{j-1}} P_{\tau_j}$ where $(P_0, P_1, P_2, P_3) = (I, X, Y, Z)$, $a_{m+1} = a_0$, $b_{m+1} = b_0 = 1$, and $P_{\tau_j} = T^{a_{j-1}} P_{b_j} P_{b_{j-1}} T^{-a_{j-1}}$.
3. Calculate $|\langle 0|G'_{m+1:1}|0\rangle|^2$

We repeat the above steps for 100 random sequences for each value of $m \in \{4, 8, 16, ..., 2048\}$ and fit the averages for each sequence to the model $\mathbb{E}_{\vec{\mathcal{C}}} = 0.5 p_{est}^m + 0.5$ using Mathematica's NonlinearModelFit function weighted by the variance over the 100 sequences at each m, where we have set $A = B = 0.5$ because of the choice of SPAM and because the noise is unital.

7 Alternative analysis of RB with gate-dependent noise

We now show how the analyses of gate-dependent noise of Refs. [12, 28, 24, 26] are too loose to justify fitting RB decay curves to a single exponential in practical regimes. We will first show that the higher-order terms in the original analysis of RB give a systematic uncertainty of the infidelity that dominates the infidelity, even when the fidelity is dominated by stochastic errors. We will then discuss subsequent analysis and how it also results in large systematic errors.

7.1 Initial RB Analysis

The initial method of analyzing RB under gate-dependent noise was to perform a perturbation expansion to perturb about a perfect twirl [12, 28] of the average noise,

$$\mathcal{E} = \mathbb{E}_{G \in \mathbb{G}}(\mathcal{G}^\dagger \tilde{\mathcal{G}}). \tag{46}$$

Substituting $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{E} + \mathcal{G}\Delta_G$ into the average map over all sequences,

$$\mathbb{E}_{\vec{G} \in \mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:1}) \tag{47}$$

and using the triangle inequality and submultiplicativity, the kth order terms in Δ_G contribute at most

$$\binom{m+1}{k} \gamma^k = \binom{m+1}{k} \mathbb{E}_{G \in \mathbb{G}}(\|\Delta_G\|_{1\to 1}) \tag{48}$$

to the average survival probability over all sequences of length m, $\langle\langle Q|\mathbb{E}_{\vec{G} \in \mathbb{G}^m}(\tilde{\mathcal{G}}_{m+1:1})|\rho\rangle\rangle$. Ref. [28] also obtained similar conditions for time-dependent noise and derived the first-order correction. As stated in Ref. [12], higher-order terms will be negligible provided $m\gamma \ll 1$, so that there is a tension between the amount of gate-dependent noise and the values of m required to obtain a reasonable fit. However, there exist noise models such that $m\gamma$ is not negligible [26].

The analysis of Ref. [28] also holds for the more general gate-dependent noise model $\mathcal{L}_G\mathcal{G}\mathcal{R}_G$, where $\mathcal{L}$ and $\mathcal{R}_G$ are gate-independent and gate-dependent noise processes acting from the left and right (that is, after and before a ideal gate) respectively, so that some optimization of γ is possible.

However, we now show that the higher-order terms cannot be neglected when fitting data in experimental regimes. Suppose the average survival probability over all sequences of some fixed length m is measured to be f_m. From the zeroth-order model with $p = 1 - 2r$,

$$r = \frac{1}{2} - \frac{1}{2}\sqrt[m]{\frac{f_m - B}{A}}, \tag{49}$$

ignoring any finite measurement statistics. However, by eq. (48), the first-order perturbations add some $\delta \in [-m\gamma, m\gamma]$ to f_m, so that the true value of r is

$$r = \frac{1}{2} - \frac{1}{2}\sqrt[m]{\frac{f_m - B - \delta}{A}} \approx \frac{1}{2} - \frac{1}{2}\sqrt[m]{\frac{f_m - B}{A}} + \frac{\delta}{2m(f_m - B)}, \tag{50}$$

where we have assumed $m \gg 1$ for the approximation, so that the first-order perturbation terms add a systematic uncertainty of $\gamma/2(f_m - B)$ to the estimate of r.

Therefore the systematic ucertainty will dominate the estimate of the infidelity unless $\gamma \approx r(\mathcal{E})$. To illustrate how impractical this requirement is, and thus the value of the tighter analysis in section 5, consider the following hypothetical implementation of the single-qubit Clifford group (that is, the 24 elements of SU(2) that permute the single-qubit Paulis X, Y and Z up to an overall sign). Supoose that half the elements are implemented with only depolarizing noise, that is, $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{D}_\nu$, and the other half are implemented with an additional rotation around the z-axis of the Bloch sphere by some angle $\theta \in (0, \frac{\pi}{2})$, that is, $\tilde{\mathcal{G}} = \mathcal{G}\mathcal{D}_\nu\mathcal{Z}_\theta$ where

$$Z_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}. \tag{51}$$

For this noise model, the average noise is $\mathcal{E} = \mathcal{D}_\nu(I_4 + \mathcal{Z}_\theta)/2$ by eq. (46) and so

$$\begin{aligned} \gamma &= \frac{1}{2}\|\mathcal{D}_\nu - \mathcal{D}_\nu(I_4 + \mathcal{Z}_\theta)/2\|_{1\to 1} + \frac{1}{2}\|\mathcal{D}_\nu\mathcal{Z}_\theta - \mathcal{D}_\nu(I_4 + \mathcal{Z}_\theta)/2\|_{1\to 1} \\ &= \frac{\nu}{2}\|\mathcal{Z}_\theta - I_4\|_{1\to 1} \\ r(\mathcal{E}) &= \frac{1}{6}\Big(3 - 2\nu - \nu\cos\theta\Big) \end{aligned} \tag{52}$$

using $1 - r(\mathcal{E}) = f(\mathcal{E}, \mathcal{I}) = 1/2 + \sum_{\sigma=X,Y,Z}\mathrm{Tr}[\sigma\mathcal{E}(\sigma)]/12$ [36]. The maximization in the induced trace norm will be achieved by any state in the xy plane (e.g., $(|0\rangle + |1\rangle)/\sqrt{2}$), giving

$$\gamma = \nu|\sin(\theta/2)|. \tag{53}$$

For the estimate based on the zeroth-order model to be valid to within a factor of two, we require $\gamma \leq r(\mathcal{E})$, that is, $|\theta| \lesssim 1-\nu$. For example, the systematic uncertainty in the estimate of r is $9r$ (i.e., under-reporting the error rate r by an order of magnitude) when $\nu = 0.99$ and gate-dependent errors account for 10% of the total infidelity ($\theta = 0.09$ in the above model) or when $\nu = 0.999$ and the gate-dependent errors account for just 1% of the total infidelity ($\theta = 0.009$ in the above model).

One might expect that the above analysis is pessimistic and that the contributions from gate-dependent perturbations at different sequence lengths would average out. However, Ref. [26] presented an explicit model wherein the contributions from gate-dependent perturbations are systematic.

7.2 Subsequent analyses of RB

Ref. [26] demonstrated that the RB decay rate could differ from the average gate infidelity of the average noise defined in eq. (46) by orders of magnitude. Ref. [26] also presented a significantly improved analysis of the original RB protocol, which can be viewed as an application of the method of [24] to the original RB protocol. However, this analys has three crucial drawbacks.

First, no interpretation of the decay parameter was provided, leading to frequent discussions of "the RB number" r_{RB} at conferences. In the original version, Ref. [26] states that "it is not clear that r (a function of the RB decay parameter) should be called (the) "fidelity", except in the special case of gate-independent error maps," although this statement has since been revised based on the analysis in the present paper.

Second, the analysis of Ref. [26] is vulnerable to the criticism it makes of Ref. [28], namely, that the infidelity $r(\mathcal{E})$ of a fixed decomposition of the noise (defined in eq. (46)) can differ from r_{RB} by orders of magnitude. This is problematic for the analysis of Ref. [26] (and also for that of Ref. [24]) because they bound a perturbation from an exponential of the form $(1-r_{RB})^m$ (ignoring small dimensional factors) by the diamond norm with respect to a fixed decomposition of the noise, giving a perturbation on the order of $\sqrt{r(\mathcal{E})} \gg r(\mathcal{E}) \gg r_{RB}$. While some gauge optimization of their bound is possible (especially using a gauge motivated by the present analysis), it is unclear whether such optimization can close the gap between $r(\mathcal{E})$ and r_{RB}. Fortunately, the present paper obviates the potential problem by providing a much sharper bound.

Finally, even if gauge optimization allows the perturbation in their method to be bounded by $\sqrt{r_{RB}}$, this still renders their analysis useless in two key regimes, namely, for "large" error rates or in the linear part of the decay.

For "large" error rates where $r_{RB} \approx 0.01$, which accounts for most RB experiments to date, the best bound that can be put on the perturbation without knowing the full experimental noise process is approximately $\sqrt{r_{RB}} = 0.1$. With such a large perturbation, any sequence lengths that have an average survival probability over 0.9 or under 0.6 are effectively useless assuming SPAM values of $A = B = 0.5$ (and the range of values that contain any information decreases with A).

In the linear part of the decay, such as in the numerics of Ref. [26], the systematic uncertainty due to the perturbation dominates the decay. For concreteness, consider the linear regime for a single qubit,

$$\mathbb{E}_{\vec{G}\in\mathbb{G}^m}(Q_{\vec{G}}) \approx A + B - 2Am[1 - p(\mathcal{RL})]. \tag{54}$$

where $p(\mathcal{RL})$ is the true RB decay parameter from theorem 4. If $p(\mathcal{RL})$ is estimated via the slope between $m = m_1$ and $m = m_2$, then the unknown perturbation contributes a systematic uncertainty of approximately $\sqrt{r(\mathcal{E})}/(m_1 - m_2)$ to the estimate of p, so that essentially no information is gained unless $m_1 - m_2 \gg \sqrt{r(\mathcal{E})}$. For the current record value of $1 - p \approx 10^{-6}$, this translates to requiring $m_2 - m_1 > 1000$ assuming $1 - p(\mathcal{RL}) \approx r(\mathcal{E})$. However, the assumption that $1 - p(\mathcal{RL}) \approx r(\mathcal{E})$ is unjustified as shown by their analysis, so that the required spacing of sequence lengths in order to overcome the systematic uncertainty due to their bound is unknown, rendering their analysis impractical. Consequently, the analysis of Ref. [26] does not explain the empirical success of RB to date or justify fitting to a single exponential without further analysis.

8 Conclusion

We have proven that randomized benchmarking experiments result in a single exponential with a negligible perturbation under gate-dependent noise. Moreover, we proved that the decay parameter is the fidelity of the average noise between idealized gates. More specifically, we can write $\tilde{\mathcal{G}} = \mathcal{LGR}_G$ where $\mathcal{L}$ is a solution to eq. (13). Then for trace-preserving noise the RB decay parameter is $\mathbb{E}_{G\in\mathbb{G}}[p(\mathcal{R}_G\mathcal{L})]$, that is, the average survival probability between ideal gates. As $\tilde{\mathcal{G}} \to \mathcal{S}^{-1}\tilde{\mathcal{G}}\mathcal{S}$ maps $\mathcal{L} \to \mathcal{SL}$ and $\mathcal{R} \to \mathcal{RS}^{-1}$ in theorem 2, the decay parameter $p = \mathbb{E}_{G\in\mathbb{G}}(\mathcal{R}_G\mathcal{L})$ is manifestly gauge invariant, answering the recent criticism of Ref. [26].

By rigorously reducing the potential impact of gate-dependent errors, the present work enables randomized benchmarking to more reliably diagnose the presence of non-Markovian noise. More precisely, there are two key assumptions that may not hold well enough for the standard exponential decay model to be valid, namely, Markovianity and time-independence. However, the only effect that time-dependent noise can have is to alter the rate at which the survival probability decays [8, 30]. Consequently, any "revivals" in RB experiments, that is, statistically significant increases in the average survival probability as m increases [23], can be attributed directly to the noise being non-Markovian.

An intriguing observation from the present analysis is that gate-dependent fluctuations may produce a significant deviation from the exponential fit for short sequences. This deviation is possible because the first few values of m are the most sensitive to gate-dependent fluctuations, which may be as large as 0.1 for $m = 1$ and 0.01 for $m = 2$. Consequently, these sequence lengths should be omitted when fitting to eq. (8), but can also indicate the presence of gate-dependent noise.

Alternative protocols based on randomized benchmarking have been developed that also assume gate-independent noise [13, 14, 15, 16, 17, 18, 19]. We leave applying the current analysis to these protocols to future work.

One main open problem of the current work is that the solutions $\mathcal{L}$ and $\mathcal{R}$ from theorem 2 are not guaranteed to be CPTP maps, even under rescaling.

Acknowledgments

The author acknowledges helpful discussions with Robin Blume-Kohout, Arnaud Carignan-Dugas, Joseph Emerson, Steve Flammia, Chris Granade, Richard Kueng, Robin Harper, Jonas Helsen, Thomas Monz, Jiaan Qi and Philipp Schindler. This research was supported by the U.S. Army Research Office through grant W911NF-14-1-0103. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund.

References

[1] Isaac L. Chuang and Michael A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, Journal of Modern Optics, **44**, 2455 (1997).

[2] J. F. Poyatos, J. Ignacioi Cirac, and P. Zoller, Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate, Physical Review Letters **78**, 390 (1997).

[3] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin, Practical Characterization of Quantum Devices without Tomography, Physical Review Letters **107**, 210404 (2011).

[4] Steven T. Flammia and Yi-Kai Liu, Direct Fidelity Estimation from Few Pauli Measurements, Physical Review Letters **106**, 230501 (2011).

[5] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert, Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators, New Journal of Physics **14**, 095022 (2012).

[6] Daniel M. Reich, Giulia Gualdi, and Christiane P. Koch, Optimal Strategies for Estimating the Average Fidelity of Quantum Gates, Physical Review Letters **111**, 200401 (2013).

[7] Martin Kliesch, Richard Kueng, Jens Eisert, and David Gross, Guaranteed recovery of quantum processes from few measurements, arXiv:1701.03135 [quant-ph].

[8] Joseph Emerson, Robert Alicki, and Karol Życzkowski, Scalable noise estimation with random unitary operators, Journal of Optics B **7**, S347 (2005).

[9] Benjamin Lévi, Cecilia C López, Joseph Emerson, and David G. Cory, Efficient error characterization in quantum information processing, Physical Review A **75**, 022314 (2007).

[10] Emanuel Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and David J. Wineland, Randomized benchmarking of quantum gates, Physical Review A **77**, 012307 (2008).

[11] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Physical Review A **80**, 012304 (2009).

[12] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Physical Review Letters **106**, 180504 (2011).

[13] Joseph Emerson, Marcus P. da Silva, Osama Moussa, Colm A. Ryan, Martin Laforest, Jonathan Baugh, David G. Cory, and Raymond Laflamme, Symmetrized characterization of noisy quantum processes. Science **317**, 1893 (2007).

[14] Easwar Magesan, Jay M. Gambetta, Blake R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, Mark B. Ketchen, and Matthias Steffen, Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking, Physical Review Letters **109**, 080505 (2012).

[15] Joel J. Wallman, Christopher Granade, Robin Harper, and Steven T. Flammia, Estimating the Coherence of Noise, New Journal of Physics **17**, 113020 (2015).

[16] Joel J. Wallman, Marie Barnhill, and Joseph Emerson, Robust Characterization of Loss Rates, Physical Review Letters **115**, 060501 (2015).

[17] Joel J. Wallman, Marie Barnhill, and Joseph Emerson, Robust characterization of leakage errors, New Journal of Physics **18**, 043021 (2016).

[18] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson, Characterizing universal gate sets via dihedral benchmarking, Physical Review A **92**, 060302(R) (2015).

[19] Andrew W. Cross, Easwar Magesan, Lev S. Bishop, John A. Smolin, and Jay M. Gambetta, Scalable randomised benchmarking of non-Clifford gates, npj Quantum Information **2**, 16012 (2016).

[20] Antonio D. Córcoles, Jay M. Gambetta, Jerry M. Chow, John A. Smolin, Matthew Ware, Joel Strand, B. L. T. Plourde, and Matthias Steffen, Process verification of two-qubit quantum gates by randomized benchmarking, Physical Review A **87**, 030301(R) (2013).

[21] R. Barends, Julian Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, Austin 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 John M. Martinis, Superconducting quantum circuits at the surface code threshold for fault tolerance. Nature **508**, 500 (2014).

[22] Julian Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, Austin G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis, Optimal Quantum Control Using Randomized Benchmarking, Physical Review Letters **112**, 240504 (2014).

[23] Jeffrey M. Epstein, Andrew W. Cross, Easwar Magesan, and Jay M. Gambetta, Investigating the limits of randomized benchmarking protocols, Physical Review A **89**, 062321 (2014).

[24] Tobias Chasseur and Frank K. Wilhelm, Complete randomized benchmarking protocol accounting for leakage errors, Physical Review A **92**, 042333 (2015).

[25] Harrison Ball, Thomas M. Stace, Steven T. Flammia, and Michael J. Biercuk, Effect of noise correlations on randomized benchmarking, Physical Review A **93**, 022303 (2016).

[26] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-kohout, What randomized benchmarking actually measures, Physical Review Letters **119**, 130502 (2017).

[27] Joshua Combes, Christopher Granade, Christopher Ferrie, and Steven T. Flammia, Logical Randomized Benchmarking, arXiv:1702.03688 [quant-ph].

[28] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson, Characterizing quantum gates via randomized benchmarking, Physical Review A **85**, 042311 (2012).

[29] Yuval R. Sanders, Joel J. Wallman, and Barry C. Sanders, Bounding quantum gate error rate based on reported average fidelity, New Journal of Physics **18**, 012002 (2016).

[30] Joel J. Wallman, and Steven T. Flammia, Randomized benchmarking with confidence, New Journal of Physics **16**, 103032 (2014).

[31] Christopher Granade, Christopher Ferrie, and David G. Cory, Accelerated randomized benchmarking, New Journal of Physics **17**, 013042 (2015).

[32] Michael A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Physics Letters A **303**, 249 (2002).

[33] F. L. Bauer and C. T. Fike, Norms and exclusion theorems, Numerische Mathematik **2**, 137 (1960).

[34] David Pérez-García, Michael M. Wolf, Denes Petz, and Mary Beth Ruskai, Contractivity of positive and trace-preserving maps under L_p norms, Journal of Mathematical Physics **47**, 083506 (2006).

[35] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Kenneth Rudinger, Jonathan Mizrahi, Kevin Fortier, and Peter Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nature Communications **8**, 14485 (2017).

[36] Mark D. Bowdrey, Daniel K. L. Oi, Anthony J. Short, Konrad Banaszek, and Jonathan A. Jones, Fidelity of single qubit maps, Physics Letters A **294**, 258 (2002).