CORRIGENDUM

Corrigendum: Randomized benchmarking with confidence (2014 New J. Phys. 16 103032)

Joel J Wallman1,2 and Steven T Flammia2

1 Institute for Quantum Computing, University of Waterloo, Waterloo, Canada
2 Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, NSW 2006, Australia

The original proof of lemma 7 in [1] contained an error due to different normalization conventions relative to a reference. Here we present an alternative, simpler proof of lemma 7. The original paper also transposed the definition of the Liouville representation, which should be $\epsilon_{j,k} = \langle A_j, \epsilon(A_k) \rangle$.

Lemma 7. Let $\Delta$ be a Hermiticity-preserving linear map between $d$-dimensional operators. Then the following inequalities bound the diamond norm and are saturated:

$$\| J(\Delta) \|_1 \leq \| \Delta \|_0 \leq d \| J(\Delta) \|_1.$$  \hspace{1cm} (1)

Proof. We first prove the lower bound and show that it is saturated. We have

$$\| \Delta \|_0 = \sup_{\psi} \| \Delta \otimes I_d(\psi) \|_1 \geq \| \Delta \otimes I_d(\Phi) \|_1 = \| J(\Delta) \|_1.$$  \hspace{1cm} (2)

To see that the above inequality is saturated, simply let $\Delta = I_d$.

To prove the upper bound, we write

$$\| \Delta \|_0 = d \sup_{\rho} \| (I_d \otimes \sqrt{\rho_0})J(\Delta)(I_d \otimes \sqrt{\rho_1}) \|_1 : \rho_0, \rho_1 \in D_d$$  \hspace{1cm} (3)

which follows from theorem 6 of [2], while being careful to note that our convention for $J(\Delta)$ differs from [2] by a factor of $d$. Using [3, proposition IV.2.4], the inequality $\| ABC \|_1 \leq \| A \|_\infty \| B \|_1 \| C \|_\infty$ together with $\| \rho \|_\infty \leq 1$ for any state $\rho$ implies that

$$\| \Delta \|_0 \leq d \sup_{\rho} \{ \| (I_d \otimes \sqrt{\rho_0})J(\Delta)(I_d \otimes \sqrt{\rho_1}) \|_1 : \rho_0, \rho_1 \in D_d \} \leq d \| J(\Delta) \|_1.$$  \hspace{1cm} (4)

To see that this bound is saturated, let $\Delta(A) = |0\rangle \langle 0|A|0\rangle \langle 0|$. \hfill $\square$

References

[1] Wallman J and Flammia S T 2016 Randomized benchmarking with confidence New J. Phys. 18 103032
[2] Watrous J 2012 Simpler semidefinite programs for completely bounded norms (arXiv:1207.5726v2)
[3] Bhatia R 1997 Matrix Analysis (Berlin: Springer)
Randomized benchmarking with confidence

Joel J Wallman\textsuperscript{1,2} and Steven T Flammia\textsuperscript{2}

\textsuperscript{1} Institute for Quantum Computing, University of Waterloo, Waterloo, Canada
\textsuperscript{2} Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, NSW 2006, Australia
E-mail: joel.j.wallman@gmail.com

Received 9 May 2014, revised 21 July 2014
Accepted for publication 8 September 2014
Published 20 October 2014

New Journal of Physics 16 (2014) 103032
doi:10.1088/1367-2630/16/10/103032

Abstract
Randomized benchmarking is a promising tool for characterizing the noise in experimental implementations of quantum systems. In this paper, we prove that the estimates produced by randomized benchmarking (both standard and interleaved) for arbitrary Markovian noise sources are remarkably precise by showing that the variance due to sampling random gate sequences is small. We discuss how to choose experimental parameters, in particular the number and lengths of random sequences, in order to characterize average gate errors with rigorous confidence bounds. We also show that randomized benchmarking can be used to reliably characterize time-dependent Markovian noise (e.g., when noise is due to a magnetic field with fluctuating strength). Moreover, we identify a necessary property for time-dependent noise that is violated by some sources of non-Markovian noise, which provides a test for non-Markovianity.

Keywords: quantum information, randomized benchmarking, quantum noise

1. Introduction

One of the key obstacles to realizing large-scale quantum computation is the need for error correction and fault tolerance [1], which require the coherent implementation of unitary operations to high precision. Characterizing the accuracy of an experimental implementation of a unitary operation is therefore an important prerequisite for constructing a large-scale quantum computer.
It is possible to completely characterize an experimental implementation of a unitary using full quantum process tomography [2, 3]. However, this approach has several major deficiencies when applied to large quantum systems. Firstly, it is provably exponential in the number of qubits of the system for any procedure that can identify general noise sources and hence it cannot be performed practically for even intermediate numbers of qubits, despite improvements such as compressed sensing [4, 5]. Secondly, it is sensitive to state preparation and measurement (SPAM) errors, which create a noise floor below which an accurate process estimation becomes impossible [6]. Finally, it does not capture any notion of systematic, time-dependent errors that can arise from applying many unitaries in sequence.

One can avoid the exponential scaling by accepting a partial characterization of an experimental implementation. A partial characterization of, for example, the average error rate and/or the worst-case error rate compared to a perfect implementation of a target unitary is typically enough to determine whether an experimental implementation of a unitary is sufficient for achieving fault-tolerance in a specific scheme for fault-tolerant quantum computation. Such partial characterizations can be obtained efficiently (in the number of quantum systems) using either randomized benchmarking [7–12] or direct fidelity estimation [13, 14].

While direct fidelity estimation gives an unconditional and assumption-free estimate of the average gate fidelity, it is prone to SPAM errors, which leads to conflation of noise sources. Thus, a key advantage of randomized benchmarking is that it is not sensitive to SPAM errors. Unfortunately, however, current proposals for randomized benchmarking assume that the noise is time-independent, although time-dependence can be partially characterized by a deviation from the expected fidelity decay curve [9, 10]. Furthermore, experimental implementations of randomized benchmarking typically use on the order of 100 random sequences of Clifford gates, which is three orders of magnitude smaller than the number of sequences suggested by the rigorous bounds in [10] to obtain an accuracy comparable to the claimed experimental accuracies [12, 15]. Numerical investigations of a variety of noise models have shown that between 10–100 random sequences for each length are sufficient to provide a tight estimate of the average gate fidelity [16]. Ideally, one would like to combine the advantages of both randomized benchmarking and direct fidelity estimation to achieve a method that is insensitive to SPAM, requires few measurements, is nearly assumption-free (i.e., does not assume a specific noise model), and comes with rigorous guarantees on the errors involved.

In this paper, we provide a new analysis of randomized benchmarking which brings it closer in line with this ideal. We first show that the standard protocol can be modified to provide a means of estimating the time-dependent average gate fidelity (which characterizes the average error rate), provided that the gate-dependent fluctuations at each time step are sufficiently small. Under the assumption that the noise is Markovian (that is, that the noise can be written as a sequence of noisy channels acting on the system of interest), all the time-dependent parameters that are estimated by our procedure are upper-bounded by 1, so if some of the parameters are observed to be greater than 1, the experimental noise must be non-Markovian.

We then provide a rigorous justification for taking a small number of random sequences at each length that is on the same order as used in practice by obtaining bounds on the variance due to sampling gate sequences. Our work complements the approach of [16], where it was shown that the width of the confidence interval for the parameters extracted from randomized benchmarking is on the order of the square root of the variance. Our work therefore proves that this confidence interval is generally very narrow, that is, the parameters extracted from randomized benchmarking are determined with high precision.
Numerically, we observe that our bounds (at least for qubits) are saturated and so cannot be improved without further assumptions on the noise (e.g., that the noise is diagonal in the Pauli basis). Therefore any experiments using fewer random sequences than justified by our analysis (unless there is solid evidence that the noise has a specific structure) will potentially underestimate the error due to sampling random sequences.

As a particular example, our results provide a rigorous proof that for single-qubit noise with an average error rate of $10^{-4}$, the error for randomized benchmarking with 100 random sequences of 100 random gates will be less than 0.9% with 99% confidence. If we use the parameters estimated in the experiment of [15], with 100 random sequences of length 987 at an average error rate of $2 \times 10^{-5}$, we find the error is less than 0.8% with 99% confidence.

We emphasize that our results are solely in terms of the number of random gate sequences, and a given sequence must still be repeated many times to gather statistics about expectation values of an observable. This is of course an unavoidable consequence of quantum mechanics. However, these statistical fluctuations in the estimates of expectation values can be analyzed separately with standard statistical tools for binomial distributions or with the recent Bayesian methods introduced in [17] and combined seamlessly with our results.

In order to give a rigorous statement of results, we will first review the randomized benchmarking protocol.

2. The randomized benchmarking protocol

The goal of randomized benchmarking is to efficiently but partially characterize the average noise in an experimental implementation of a group $G = \{g_1, \ldots, g_9\} \subset U(d)$ of operations acting on a $d$-dimensional quantum system. In order to characterize the average noise in an implementation of $G$ using randomized benchmarking, we require $G$ to be a unitary two-design (e.g., the Clifford group on $n$ qubits for $d = 2^n$), meaning that sampling over $G$ reproduces the second moments of the Haar measure [18, 19]. To accomplish this, the following protocol is implemented.

- Choose a random sequence $s = s_1 \ldots s_m \in \mathbb{N}_{|G|}^m$ of $m$ integers chosen uniformly at random from $\mathbb{N}_{|G|} = \{1, \ldots, |G|\}$.
- Prepare a $d$-dimensional system in some state $\rho$ (usually taken to be the pure state $|0\rangle$).
- At each time step $t = 0, \ldots, m$, apply $g_t$ where $g_t = g_{s_t}$ and $g_0 = \prod_{t=1}^m g_t^{-1}$. Alternatively, to perform interleaved randomized benchmarking for the gate $g_{\text{int}} \in G$, apply $g_{t,\text{int}}$ where $g_{t,\text{int}} = g_{s_t} g_t$ for $t \neq 0$ and, as before, $g_{0,\text{int}} = \prod_{t=1}^m g_t^{-1}$. (In general, each gate must be compiled into a sequence of elementary gates as well.)
- Perform a POVM $\{E, 1 - E\}$ for some $E$ (usually taken to be $|0\rangle\langle 0|$) and repeat with the sequence $s$ sufficiently many times to obtain an estimate of the probability $F_{m,s} = p(E|s, \rho)$ to a suitable precision.

We can regard the probability $F_{m,s}$ as a realization of a random variable $F_m$. We will denote the variance of the distribution $\{F_{m,s} : s \in \mathbb{N}_{|G|}^m\}$ for a fixed $m$ by $\sigma_m$. Averaging $F_{m,s}$ over a number of random sequences will give an estimate $\hat{F}_m$ of $F_m$, the average of $F_{m,s}$ over all sequences $s$ of fixed length $m$ (that is, $\hat{F}_m$ is the expectation of the random variable $F_{m,s}$). The accuracy of this estimate will be a function of the number of random sequences and $\sigma_m^2$. 

New J. Phys. 16 (2014) 103032 J J Wallman and S T Flammia
Obtaining estimates $\hat{F}_m$ for multiple $m$ and fitting to the model

$$F_m = A + Bf^m$$

will give an estimate of $f$ provided that the noise does not depend too strongly on the target gate [10], where [20]

$$f = \frac{dF_{\text{avg}}(E) - 1}{d - 1}$$

and

$$F_{\text{avg}}(E) = \int d\psi \text{ Tr} [\psi E(\psi)]$$

is the average gate fidelity of a noise channel $E$ with respect to the identity channel and $d\psi$ is the uniform Haar measure over all pure states. The average gate fidelity of $E$ gives the average probability that preparing a state $\psi$, applying $E$ and then measuring $\{\psi, 1 - \psi\}$ will give the outcome $\psi$, averaged over all pure states $\psi$.

For standard randomized benchmarking, $E$ is the error channel per operation, averaged over all operations in $G$. For interleaved benchmarking, $E$ is the error channel on a composite channel, namely, the interleaved channel composed with an element of $G$, averaged over all $G$. We note in passing that separating the error in the interleaved channel from the error in the composite channel is one of the key difficulties in obtaining meaningful results from interleaved benchmarking [21], though we do not address this issue here.

3. Statement of results and paper outline

The first principal contribution of this paper is to show that the number of random sequences that need to be averaged is comparable to the number actually used in contemporary experiments (compared to previous best estimates, which require three orders of magnitude more random sequences than currently used). The second principal contribution is to show that randomized benchmarking can be used to characterize time-dependent fluctuations in the noise strength.

In more detail, and in order of appearance, we show the following.

- We use the results derived later in the paper to obtain explicit confidence intervals for the estimates $\hat{F}_m$ when $mr \ll 1$, where $r = 1 - F_{\text{avg}}(E)$ is the average gate infidelity (section 4.1).
- Again, using results derived later, we show that a more thorough analysis of randomized benchmarking data can be used to characterize time-dependent Markovian noise, and consequently as a sufficient condition for the presence of non-Markovian noise in a system (section 4.2).
- We review representation theory and the Liouville representation of quantum channels and prove some elementary results (section 5). We give an explicit proof of bounds on the diamond norm (which characterizes the worst-case error rate) in terms of the average gate fidelity (which characterizes the average error rate). These give slight improvements over previously stated (but unproven) bounds (section 5.4).
We derive an expression for the mean of the randomized benchmarking distribution with time-dependent noise (section 6.1).

We show that the variance for randomized benchmarking \( d \)-level systems with average gate infidelity and sequences of length \( m \) satisfies

\[
\sigma_m^2 \leq 4d(d+1)mr + O(m^2r^2d^4).
\]

Furthermore, we provide an argument that suggests that this bound can be improved to

\[
\sigma_m^2 \leq mr + O(m^2r^2d^4).
\]

For qubits, we improve the upper bound to

\[
\sigma_m^2 \leq m^2r^2 + \frac{7mr^2}{4} + 6\delta mr + O(m^2r^3) + O(\delta m^2r^2),
\]

where \( \delta \) quantifies the deviation from preparations and measurements in a Pauli eigenstate. We use this improved bound to derive confidence intervals that rigorously justify the use of a small number of random sequences for qubits in the regime \( mr \ll 1 \).

For the special case of single-qubit noise that is diagonal in the Pauli basis, we further improve the upper bound to

\[
\sigma_m^2 \leq \frac{11mr^2}{4} + O(m^2r^3),
\]

which is independent of preparations and measurements.

We show that the variance for unital (but nonunitary) channels decays exponentially to zero asymptotically, while the variance for nonunital noise converges exponentially to a positive constant proportional to the degree of nonunitality (as suitably quantified).

We prove that our results are robust under gate-dependent noise, which is one of the key assumptions under which randomized benchmarking produces a meaningful result. Furthermore, since our results apply to interleaved randomized benchmarking, gate dependence can be experimentally tested and used to bound the contribution from gate-dependent terms.

### 4. Analyzing data from randomized benchmarking with finite sampling

In this section, we summarize the implications of our results for analyzing the data obtained from randomized benchmarking experiments. In particular, we derive confidence intervals for the estimates \( \hat{F}_m \) of \( \bar{F}_m \) and show how randomized benchmarking can be used to characterize time-dependent noise.

#### 4.1. Confidence interval for randomized benchmarking

For a fixed sequence length \( m \), randomized benchmarking provides an estimate \( \hat{F}_m \) of \( F_m \), which is exact in the limit when all random sequences are sampled. We will only consider the variance \( \sigma_m^2 \) due to sampling a finite number \( K_m \) of random sequences of length \( m \), and we ignore the random fluctuations resulting from the use of a finite number of measurements to estimate a probability.
In [10], the variance-independent form of Hoeffding’s inequality was used to estimate the number of sequences $K_m$ required to obtain a given level of accuracy. The estimate in [10] erroneously restricted the range of the random variable in Hoeffding’s inequality. That is, they assumed that all the probabilities $F_{ms}$ lay in a strict subset of $[0, 1]$. This assumption, while valid for depolarizing noise, is not valid in general. A simple counterexample is where the noise is a single-qubit preparation channel into the $\ket{0}\bra{0}$ state and $\rho = E = \ket{0}\bra{0}$. Then any sequence of $m$ gates ending in an identity gate or a $z$-axis rotation has $F_{m,s} = 1$, while any sequence ending in an $X$ gate gives $F_{m,s} = 0$. Correcting for this (which does not change any of the conclusions of [10]), the variance-independent form of Hoeffding’s inequality requires $10^5$ samples to ensure that the estimate $\tilde{F}_m$ is within $5 \times 10^{-3}$ of the true mean $\tilde{F}_m$ with 99% probability. However, many experimental implementations of randomized benchmarking only use 30–100 sequences for each value of $m$ [12, 15, 22].

One of the principal contributions of this paper is to provide a theoretical justification for choosing a relatively small number of sequences by showing that the variance is small for the short sequences that are of practical relevance. For the special case of qubits, we show that even for small $m$ (e.g. $m \approx 100$) the variance is at most $4 \times 10^{-4}$ for currently achievable gate infidelities $r \approx 10^{-4}$, which is comparable to the numerical estimates presented in figure 1. Utilizing this very small variance gives substantial improvements over the previous rigorous

---

**Figure 1.** Plot of a random sampling of the exact variance $\sigma^2_m$ as a function of the sequence length $m$ for randomized benchmarking with 100 randomly generated, time-independent noisy qubit channels. The noise was sampled from the set of extremal qubit channels, characterized in [24], with average gate infidelity $r \lesssim 2.69 \times 10^{-4}$. Each channel was evolved for increasing sequence lengths to track the behavior of the variance as a function of $m$, which is why the data points track parabolic curves (furthermore, the spread in the parabolic curves is generated by the spread in the infidelity of the samples). The green curve is the upper bound $\sigma^2_m = m^2 r^2 + \frac{7mr^2}{4}$, where we neglect the corrections to the bound at order $O(m^2 r^3)$ and corrections due to measurement imprecision. Note that our bound is almost optimal. Our analytic results show that the variance $\sigma^2_m$ will increase with $m$, at least until some threshold sequence length where the exponential decay for generic channels proven in theorem 17 begins to dominate.
bounds obtained in [10, 21]. However, our bound on the variance (which is numerically almost optimal for qubits) implies that $K_m$ should scale quadratically with $m$ to make the variance is independent of $m$.

Our upper bound $\sigma_m^2 \leq m^2 \epsilon^2 + \frac{7}{2} mr^2 + O(m^2 r^3)$ (for qubits, neglecting the negligible $\delta r$ terms) can be used together with a stronger version of Hoeffding’s inequality [23] to obtain a rigorous confidence interval comparable to the standard errors of the mean reported in current experiments [15]. The stronger version of Hoeffding’s inequality implies that

\[
\Pr\left( |\hat{F}_m - F_m| > \epsilon \right) \leq 2 \left[ H\left(\epsilon, \sigma_m^2\right)\right]^K,
\]

where $K$ is the number of randomly sampled sequences of length $m$ and

\[
H(\epsilon, v) = \left( \frac{1}{1 - \epsilon} \right)^{\frac{1}{1 + v}} \left( \frac{v}{v + \epsilon} \right)^{\frac{1}{1 + v}}.
\]

Consequently, sampling

\[
K = -\frac{\log\left(\frac{2}{\delta}\right)}{\log\left(H\left(\epsilon, \sigma_m^2\right)\right)}
\]

random sequences is sufficient to obtain an absolute precision of $\epsilon$ with probability $1 - \delta$. Since $r$ is determined by the fitting procedure, which in turn depends on the uncertainties, this procedure would be applied recursively with an initial upper bound on $r$. Similarly, for qudits, a straightforward generalization of the above argument can be used, but with $\sigma_m^2 \leq 4d (d + 1) mr + O(d^4 m^2 r^3)$. (There are various inefficiencies in this estimate which mean that it does not reduce to the same answer as above for $d = 2$; see theorem 10 for more details.)

To get a feel for the sort of estimates that this bound provides, consider the following parameters for a single-qubit benchmarking experiment: $m = 100$, $r = 10^{-4}$, $\epsilon = 1\%$, $\delta = 1\%$, and use our upper bound of $\sigma_m^2 = m^2 \epsilon^2 + \frac{7}{2} mr^2$ (ignoring the higher-order terms). Then our bound shows that $K = 145$ random sequences suffices. This is an improvement by orders of magnitude over the previous best rigorously justifiable upper bound of $10^5$ using the variance-independent Hoeffding inequality [10].

Importantly, however, we note that the quadratic scaling with $m$ in the regime $mr \ll 1$ seems to be necessary (see figure 1). Even in the optimal case of noise that is diagonal in the Pauli basis, $K_m$ would still need to scale linearly with $m$ to make the variance independent of $m$ (where having the variance depend on $m$ would generally cause less weight to be assigned to larger $m$ when fitting). The linear scaling can be understood intuitively as following from the fact that there are $m$ places for an error in a sequence of length $m$, and the errors could add up in the worst case. Therefore, a corollary of our result is that longer sequence lengths should be averaged over more random sequences in this regime.

Furthermore, we prove in section 7 that there are noise sources such that the variance due to sampling random sequences is constant (or decays on an arbitrarily long timescale). If such noise sources (including nonunital noise and any unitary noise, such as over- and under-rotations) are believed to be present, substantially more random sequences need to be sampled. As such, randomized benchmarking is most reliable in the regime $mr \ll 1$, although, since the next lowest order terms in our bound are $\delta m^2 r^2$ and $m^2 r^3$, the lowest order bounds on the variance should be approximately valid for $mr \approx 0.1$. 

New J. Phys. 16 (2014) 103032 J J Wallman and S T Flammia
4.2. Characterizing time-dependent noise

The original presentation of randomized benchmarking assumed that the noise was approximately time independent (i.e., independent of the time step at which the gate is applied), with any Markovian time-dependence being partially characterized by deviations from the time-independent case [10]. However, in many practical applications there may be a nonnegligible time dependence, which it would be desirable to characterize more fully.

We show that randomized benchmarking can also be used to characterize time-dependent noise, provided the gate-dependence is negligible (in the sense established in theorem 18) and that the time-dependent noise is identically distributed between different experiments. However, the number of random sequences of length \( m \) will typically need to be increased relative to the number required for time-independent noise. In particular, we will show in theorem 8 that

\[
\prod_{t=1}^{m} f_t = A + B \prod_{t=1}^{m} f_t, \tag{11}
\]

where \( A \) and \( B \) are constants that depend only upon the preparation and measurement procedures (and so account for SPAM) and the average gate fidelity at time \( t \) is \( f_t = f_t + (1 - f_t)/d \), where \( d \) is the dimensionality of the system being benchmarked. In the case of time-independent noise, \( f_t \) is a constant and equation (11) reduces to the standard equation for the gate fidelity decay curve.

By performing randomized benchmarking for a set of sequence lengths \( m_1 \) and \( m_2 \), we can estimate \( \hat{F}_{m_j} \) with associated uncertainties \( \delta_j \). Combining these estimates with a procedure for obtaining an estimate \( \hat{A} \) of \( A \) with associated uncertainty \( \delta_A \) [25], we can estimate the ratio

\[
\frac{\hat{F}_{m_2} - \hat{A}}{\hat{F}_{m_1} - \hat{A}} = \prod_{t=m_1+1}^{m_2} f_t, \tag{12}
\]

with uncertainty on the order of

\[
\delta_{1,2,A} \approx \sqrt{\left(\delta_1 + \delta_A\right)^2 + \left(\delta_2 + \delta_A\right)^2}. \tag{13}
\]

Therefore we can estimate the average gate infidelity \( r \) over the time interval \([m_1 + 1, m_2]\). From our rigorous analysis, we can infer that \( \delta_1 \) and \( \delta_2 \) will be small for small \( m_j r \), while \( \delta_A \) will be determined only by finite measurement statistics. Furthermore, when \( m_2 \approx m_1 \), \( \prod_{t=m_1+1}^{m_2} f_t \approx 1 - r (m_2 - m_1) \) and so the above method gives a reliable method of characterizing the time-dependent gate fidelity.

We also note that if there are no temporal correlations in the noise, than all of the parameters \( r_t \) (where \( r_t \) is the average gate infidelity at time \( t \)) are lower-bounded by zero. Therefore any negative values (or average values) of \( r_t \) are an indicator of temporal correlations in the noise, that is, of non-Markovian behavior.

5. Mathematical preliminaries

Randomized benchmarking involves composing random sequences of quantum channels that are sampled in a way which approximates a group average. For this reason, it is natural to consider both the representation theory of groups and the structure of quantum channels, especially the composition of channels. In this section we collect several mathematical results in
this vein that we will need to prove our main results. We begin by considering group representation theory, and in particular prove a proposition showing how the tensor product of certain representations couple together. Most of this material is standard and can be found in any textbook on the subject, e.g. [26].

5.1. Representation theory and some useful lemmas

A representation (rep) of a group \( \mathcal{G} \) is a pair \( \phi, V \), where \( V \) is vector space known as the representation space (which we always take to be \( \mathbb{R}^d \) or \( \mathbb{C}^d \) for different values of \( d \)) and \( \phi: \mathcal{G} \to \text{GL}(V) \)—where GL(V) is the general linear group over \( V \)—is a homomorphism. A rep is faithful if \( \phi \) is injective and unitary (respectively orthogonal) if \( \phi(g) \) is a unitary (respectively orthogonal) operator for all \( g \in \mathcal{G} \). The dimension of a rep is the dimension of \( V \). A subrepresentation (subrep) is a pair \( \phi_W, W \) such that \( \phi \subseteq gW \) for all \( g \in \mathcal{G} \) and \( \phi_W \) denotes the restriction of \( \phi \) to the subspace \( W \). We sometimes refer to a space, subspace, or homomorphism as being a rep or subrep, with the complementary ingredients understood from the context.

A rep is called irreducible or an irrep if the only subreps are \( \emptyset \) and \( V \). Since the reps we consider are unitary reps of compact groups, if \( W \) is a subrep of \( V \) then the orthogonal complement \( W^\perp \) is a subrep as well. Therefore any rep can be decomposed into a direct sum of irreps, which may occur with some multiplicity. Any basis that decomposes a rep into a direct sum of irreps is called a Schur basis.

The simplest rep is the trivial rep \( \mathbb{C}(1, \cdot) \), which is also an irrep. The trivial rep is defined for any group \( \mathcal{G} \) and take any element of \( \mathcal{G} \) to 1. While the trivial rep deserves its name, it frequently appears as a subrep of tensor powers of other reps and so will appear throughout this paper.

The randomized benchmarking protocol is designed so that the sequence of operators applied to a system correspond to noise channels conjugated by uniformly random elements of a group \( \mathcal{G} \). Given a rep \( \phi, V \) of a group \( \mathcal{G} \), a matrix \( A \in \text{GL}(V) \) and an element \( g \in \mathcal{G} \), we define \( A^g = \phi(g)A\phi(g^{-1}) \). The uniform average of this action on \( A \) is called the \( \mathcal{G} \)-twirl of \( A \), and is given by \( A^\mathcal{G} = |\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} A^g \).

Note that, for notational convenience, the map \( \phi \) is left implicit but will always be obvious given the dimensionality of the matrix being twirled. An important property of \( A^\mathcal{G} \) is that it commutes with the action of \( \mathcal{G} \) for any rep \( \phi, V \) (reducible or not) since \( \phi \) is a homomorphism and \( \mathcal{G} \) is a group. That is, \( A^\mathcal{G} = (A^g)^\mathcal{G} = (A^\mathcal{G})^g \) for all \( g \in \mathcal{G} \).

Expressions for the expected value \( \bar{F}_m \) and variance \( \sigma^2_m \) for the randomized benchmarking protocol for a fixed value of \( m \) will be obtained using the following propositions.

**Proposition 1.** Let \( \phi, \mathbb{C}^d \) be a nontrivial \( d \)-dimensional irreducible representation of a group \( \mathcal{G} \) and \( A \in \text{GL}(\mathbb{C}^d), B \in \text{GL}(\mathbb{C}^{d+1}) \). Then

- \( A^\mathcal{G} = a1_d \);
- \( B^\mathcal{G} = B_{11} \oplus b1_d \) (where the representation of \( \mathcal{G} \) is \( (1 \oplus \phi, \mathbb{C}^{d+1}) \)); and
- \( \sum_{g \in \mathcal{G}} \phi(g) = 0 \),

where \( a = \text{Tr} A/d \) and \( b = \left( \text{Tr} B - B_{11}\right)/d \).

**Proof.** All three statements follow directly from Schur’s lemma [26].
Proposition 2. If \((\phi, V)\) is an irreducible representation of a finite group \(G\) with a real-valued character \(\chi_{\phi}\), then the trivial representation is a subrepresentation of \((\phi, V)^{\otimes 2}\) with multiplicity 1.

Proof. As the rep is irreducible, Schur’s orthogonality relations \([26]\) give

\[
|G| = \sum_{g \in G} \chi_{\phi}(g) \chi_{\phi}(g) = \sum_{g \in G} \chi_{\phi}(g)^2 = \sum_{g \in G} \chi_{\phi^{\otimes 2}}(g) \chi_{\text{tr}}(g),
\]

where we have used \(\chi_{\phi^{\otimes 2}}(g) = \left[\chi_{\phi}(g)\right]^2\) and that the character for the trivial representation is \(\chi_{\text{tr}}(g) = 1\) for all \(g \in G\).

5.2. The Liouville representation of quantum channels

A quantum channel is a linear map \(\mathcal{E} : D_d \to D_d\), where \(D_d\) is the set of \(d\)-dimensional density operators. Quantum channels can be represented in a variety of equivalent ways, with different representations naturally suited to particular applications.

In this paper, we will primarily use the Liouville representation because it is defined so that quantum channels compose under matrix multiplication. We occasionally also use the Choi representation in order to apply results from the literature, but we will introduce it only as required.

5.2.1. States and measurements. We begin by introducing the Liouville representation (also called the transfer matrix representation) of quantum states and measurements. States and measurement effects (i.e., elements of a positive-operator valued measure, or POVM) can be viewed as channels from \(\mathcal{E} : \mathbb{R} \to D_d\) and \(\mathcal{E} : D_d \to \mathbb{R}\) respectively, hence they can be treated on the same footing as any other quantum channel. However, we introduce them separately for pedagogical and notational clarity.

In the standard formulation of quantum mechanics in terms of density operators and POVMs, a quantum state \(\rho \in D_d\) is any Hermitian, positive semi-definite operator such that \(\text{Tr} \rho = 1\). In addition, we always have \(\text{Tr} \rho^2 \in [0, 1]\). We can always choose a basis \(\mathcal{A} = \{A_0, A_1, \ldots, A_{d^2-1}\}\) of orthonormal operators for \(\text{GL}(\mathbb{C}^d)\), where orthonormality is according to the Hilbert–Schmidt inner product \(\langle A, B \rangle = \text{Tr}(A^\dagger B)\). We can expand any density operator relative to such a basis as \(\rho = \sum_j \rho_j A_j\), where \(\rho_j = \langle A_j, \rho \rangle\). Throughout this paper we set \(A_0 = 1/\sqrt{d}\), which fixes \(\rho_0 = \text{Tr} \rho / \sqrt{d}\), and makes all other \(A_j\) for \(j \neq 0\) traceless.

We can then identify a density operator \(\rho\) with a corresponding column vector

\[
|\rho\rangle = \left(\rho_0 \over \rho_j \right) \in \mathbb{C}^{d^2}
\]

such that \(\rho_j = \rho_j\) for \(j = 1, \ldots, d^2 - 1\). Here we make the important distinction between the density operator itself, \(\rho\), and the representation of \(\rho\) in terms of the column vector \(|\rho\rangle\). Note that \(|\rho\rangle\) is just a generalized version of a Bloch vector (with a different normalization) for \(d \geq 2\).

The conditions for \(\rho\) to correspond to a density operator now translate into geometric conditions on \(|\rho\rangle\). In particular, we will use the fact that \(|||\rho\rangle||^2 = \text{Tr} \rho^2 \in [0, 1]\), where \(||v||_2\) for \(v \in \mathbb{C}^{d^2}\) is the standard isotropic Euclidean norm.
Measurements in the standard formulation correspond to POVMs, that is, to sets of Hermitian, positive semidefinite operators \( \{ E_j \} \) such that \( \sum_j E_j = 1 \). As with quantum states, we can expand an element \( E \) of a POVM (an effect) as \( E = \sum_j E_j A_j^\dagger \), where \( E_j = \langle E, A_j \rangle \). We then identify an effect \( E \) with a row vector 
\[
(E) = \left( E_0 \ E \right) \in \mathbb{C}^{d^2},
\]
which must satisfy similar conditions to \( |\rho\rangle\).

In this formalism, the probability of observing an effect \( E \) given that the quantum state \( \rho \) was prepared is 
\[
p(E|\rho) = \text{Tr}(E\rho) = (E|\rho).
\]

### 5.2.2. Transformations

For simplicity, we will only consider quantum channels that are either states, measurements or completely positive and trace-preserving (CPTP) maps \( \mathcal{E} : \mathcal{D}_d \to \mathcal{D}_d \). We do not consider channels that reduce the trace or change the dimension because, while conceptually no more difficult, they require cumbersome additional notation and we do not use any such channels.

A quantum channel \( \mathcal{E} \) maps a density operator \( \rho \) to another density operator \( \mathcal{E}(\rho) \). We want to determine the map \( \mathcal{E} \) between the corresponding vectors \( |\rho\rangle \) and \( |\mathcal{E}(\rho)\rangle \). Since quantum channels are linear
\[
\mathcal{E}(\rho) = \sum_j \mathcal{E}(A_j)\rho_j,
\]
which implies that
\[
|\mathcal{E}(\rho)\rangle_k = \sum_j \langle A_k, \mathcal{E}(A_j) \rangle |\rho_j\rangle.
\]
That is, \( |\mathcal{E}(\rho)\rangle = \mathcal{E}|\rho\rangle \) where we abuse notation slightly and define \( \mathcal{E} \) as the matrix such that \( \mathcal{E}_{j,k} = \langle A_k, \mathcal{E}(A_j) \rangle \). That is, we use \( \mathcal{E} \) to denote both the abstract operator as well as its representation as a matrix acting on vectors \( |\rho\rangle \). In this representation, the identity channel is represented by \( 1_d \), the composition of two channels is given by matrix multiplication, and furthermore, the conjugate channel of a unitary channel \( \mathcal{E} \) is given by \( \mathcal{E}^\dagger \). In particular, these properties imply that the Liouville representation of the unitary channels is a faithful and unitary representation of \( \mathbb{U}(d) \) (though technically, it is a projective representation since a global phase is lost). Given our choice of \( \mathcal{A} \) (recall that we have fixed \( A_0 \)) and the fact that we consider only trace-preserving channels, we will always write the matrix representation of a quantum channel as
\[
\mathcal{E} = \begin{pmatrix} 1 & 0 \\ \alpha(\mathcal{E}) & \varphi(\mathcal{E}) \end{pmatrix}.
\]
A channel \( \mathcal{E} \) is unital (i.e., the identity is a fixed point of the channel) iff \( \alpha(\mathcal{E}) = 0 \). Therefore we can regard \( \|\alpha(\mathcal{E})\| \) as quantifying the nonunitality of \( \mathcal{E} \).

The representation \( (\varphi, \mathbb{C}^{d^2-1}) \) of \( \mathbb{U}(d) \) is irreducible [19], which will play a crucial role in our analysis of randomized benchmarking since it allows us to use tools from representation theory such as Schur’s lemma. Note also that the representation \( (\varphi, \mathbb{C}^{d^2-1}) \) of any subgroup \( \mathcal{G} \subseteq \mathbb{U}(d) \) that is a unitary two-design is also irreducible by the same argument (which can be regarded as a defining property of a unitary 2-design [19]). Therefore we can also use tools from
representation theory when considering channels twirled over a unitary two-design. This fact allows randomized benchmarking to be performed efficiently because unitary two-designs can be efficiently sampled while the full unitary group cannot [7, 18).

The representation $\varphi_g$ of $G$ will be one of the basic tools we use in this paper. As such, whenever $g$ appears in a matrix multiplication, it will refer to $\varphi_g$.

Randomized benchmarking will allow for the estimation of

$$f(E) := \frac{1}{d^2-1} \text{Tr} \varphi(E),$$

which corresponds to the average gate fidelity of $E$ with the identity channel. We will sometimes omit the argument of $\alpha$, $\varphi$ and $f$, or indicate the argument via a subscript. However, in all cases the argument will be clear from the context.

5.2.3. Properties of channels in the Liouville representation. Since the Liouville representation associates a unique matrix to each channel, we can characterize properties of quantum channels by properties of the corresponding matrix. In particular, we will consider the spectral radius

$$\rho(M) = \max_j |\eta_j(M)|,$$

and the spectral norm, $\|M\|_{\infty} = \max \sigma_j(M)$, where $\{\eta_j(M)\}$ and $\{\sigma_j(M)\}$ are the eigenvalues and singular values of a matrix $M$ respectively. These norms satisfy $\rho(M) \leq \|M\|_{\infty}$, which we will use to obtain bounds on valid quantum channels.

Proposition 3. Let $E$ be a completely positive map. Then the adjoint channel $E^\dagger$ is also completely positive.

Proof. Any map can be written as

$$E = \sum K_j \otimes L_j^T,$$

where the superscript $T$ denotes the transpose and the $K_j$ and $L_j$ are Kraus operators for $E$. We then have

$$E^\dagger = \sum K_j^\dagger \otimes L_j^{*},$$

where the * denotes complex conjugation. By Choi’s theorem on completely positive maps, $E$ is completely positive if and only if Kraus operators can be chosen so that $L_j = K_j^T$. Therefore Kraus operators $K_j^\dagger$ and $L_j^\dagger$ for $E^\dagger$ can be chosen so that $L_j^\dagger = (K_j^\dagger)^\dagger$.

Corollary 4. The adjoint channel of a unital, CPTP channel is also a unital, CPTP channel.

Proposition 5. Any CPTP channel $E: D_d \rightarrow D_d$ satisfies the following relations

(i) $\det E \leq 1$,

(ii) $\|E\|_{\infty} \leq \sqrt{d}$,

(iii) $\rho(E) \leq 1$,

(iv) $\|\alpha(E)\|_2 \leq \sqrt{d-1}.$

(21)
Inequality (i) is saturated if and only if $\mathcal{E}$ is unitary. Furthermore, if $\mathcal{E}$ is unital, then (ii) can be improved to $\|\mathcal{E}\|_\infty = 1$.

**Proof.**

(i) See [27, theorem 2].

(ii) See [28, theorem 2.1], noting that $\|\mathcal{E}\|_\infty = \|\mathcal{E}\|_{2\rightarrow 2}$.

(iii) See [29].

(iv) For any density operator $\rho$, we have $\text{Tr} \rho^2 \in [0, 1]$. In particular consider $\mathcal{E}(\mathbb{I}/d)$, which must be a density operator since $\mathbb{I}/d$ is a density operator and $\mathcal{E}$ is a quantum channel. Then

$$\text{Tr} \mathcal{E}(\mathbb{I}/d)^2 = \|\mathcal{E}(\mathbb{I}/d)\|_2^2 = \frac{1 + \|\alpha(\mathcal{E})\|_2^2}{d} \leq 1$$

(22)

gives the desired result.

Now let $\mathcal{E}$ be a unital channel. Then $\mathcal{E}^\dagger$ and thus $\mathcal{E}^\dagger \mathcal{E}$ are also channels by corollary 4. Substituting (i) into the equality $\rho(\mathcal{E}^\dagger \mathcal{E}) = \|\mathcal{E}\|_\infty$ gives the improved bound.

5.3. **Representing noisy channels**

An attempt to physically implement a quantum channel $\mathcal{E}$ will generally result in some other channel $\mathcal{E}'$, with the aim being, loosely speaking, to make $\mathcal{E}'$ as close to $\mathcal{E}$ as possible. We will now outline how noisy channels can be related to the intended channel in the linear representation.

Consider an attempt to implement a target unitary channel $\mathcal{U}$ that results in some (noisy) channel $\mathcal{E}$. Then since $\mathcal{E}$ is a real square matrix, it can be written as $\mathcal{E} = LQ$, where $L$ is a lower triangular matrix and $Q$ is an orthogonal matrix. Since $\mathcal{U}$ is an orthogonal matrix, we can always write $\mathcal{E} = \Lambda^{\text{post}}(\mathcal{U}) \mathcal{U}$, where $\Lambda^{\text{post}}(\mathcal{U}) = LQU^T$. Similarly, we can always write $\mathcal{E} = \mathcal{U}\Lambda^{\text{pre}}(\mathcal{U})$.

While the difference between these expressions is trivial for any single channel, it can cause confusion when comparing channels. Since $\Lambda^{\text{pre}}(\mathcal{U}) = \mathcal{U}^T\Lambda^{\text{post}}(\mathcal{U})\mathcal{U}$, the notion of ‘the’ noise in an implementation of $\mathcal{U}$ depends on which expression is used. (We will fix a representation below to avoid this ambiguity.)

The convergence of randomized benchmarking depends crucially upon the assumption that the noise is approximately independent of the target. However, in a general scenario, at most one of $\Lambda^{\text{post}}$ or $\Lambda^{\text{pre}}$ will be approximately independent of the target, with the specific choice depending upon the physical implementation. As a specific example, consider amplitude damping for a single qubit, which can be written as

$$\Delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{g} & 0 & 0 \\ 0 & 0 & \sqrt{g} & 0 \\ 1 - g & 0 & 0 & g \end{pmatrix}$$

(23)

in the Pauli basis $\frac{1}{\sqrt{2}}(\mathbb{I}, X, Y, Z)$, where $g \in [0, 1]$ determines the strength of the damping. Assume that this noise is applied independently from the left (i.e., $\Lambda^{\text{post}} = \Delta$), independently of the target. Then for $X$ and $Z$
\[ \Lambda^{\text{pre}}(Z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{g} & 0 & 0 \\ 0 & 0 & \sqrt{g} & 0 \\ (1 - g) & 0 & 0 & g \end{pmatrix}, \quad \Lambda^{\text{pre}}(X) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{g} & 0 & 0 \\ 0 & 0 & \sqrt{g} & 0 \\ (1 - g) & 0 & 0 & g \end{pmatrix} \] (24)

which is only independent of the target when \( g = 1 \) (i.e., when there is no noise).

In this work, we write noise operators as \textit{pre-multiplying} the target rather than post-multiplying the target as in [9]. The reason for this change is so that the residual noise term that is not averaged is in the first time step rather than the last and so is independent of the sequence length. While this simplifies the analysis, all the results of this paper can be derived for the other form with small modifications.

5.4. Measures of noise

The fidelity and trace distance between two quantum states are defined as

\[ F(\rho, \sigma) = \left\| \sqrt{\rho} \sqrt{\sigma} \right\|_1^2 \]
\[ D(\rho, \sigma) = \frac{1}{2} \| \rho - \sigma \|_1, \] (25)

respectively\(^3\), where the one-norm (or trace norm) is given by \( \| X \|_1 = \text{Tr} \sqrt{X^\dagger X} \). These two quantities are related by the Fuchs–van de Graaf inequalities [30]

\[ 1 - \sqrt{F(\rho, \sigma)} \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)}, \] (26)

where the right-hand inequality is always saturated when both states are pure. When one of the states is a pure state \( \psi \), the left-hand inequality in equation (26) can be sharpened to

\[ 1 - F(\psi, \sigma) \leq D(\psi, \sigma). \] (27)

Both of these quantities for quantum states can be promoted to distance measures for quantum channels [31]. Two such measures are the average gate fidelity and the diamond distance.

The average gate fidelity between a channel \( \mathcal{E} \) and a unitary \( \mathcal{U} \) is defined to be

\[ F_{\text{avg}}(\mathcal{E}, \mathcal{U}) = \int d\psi F[\mathcal{E}(\psi), \mathcal{U}(\psi)], \] (28)

where \( d\psi \) is the unitarily invariant Haar measure. For convenience, it is typical to define a single argument version

\[ F_{\text{avg}}(\mathcal{U}^\dagger \mathcal{E}) = F_{\text{avg}}(\mathcal{E}, \mathcal{U}), \] (29)

which is technically the average gate fidelity between \( \mathcal{U}^\dagger \mathcal{E} \) and \( \mathbb{1} \).

The \textit{diamond distance} between two quantum channels \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \) with \( \mathcal{E}_j : D_d \rightarrow D_d \) is defined in terms of a norm of their difference \( \Delta = \mathcal{E}_1 - \mathcal{E}_2 \) as follows

\[ \frac{1}{2} \| \Delta \|_o = \frac{1}{2} \sup_{\psi} \| I_d \otimes \Delta(\psi) \|_1. \] (30)

\(^3\) Note that some authors define fidelity to be the square root of the fidelity defined here.
The norm in the above definition is indeed a valid norm, called the diamond norm, and it extends naturally to any hermiticity-preserving linear map between operators. The factor of $1/2$ is to ensure that the diamond distance between two channels is bounded between 0 and 1.

The diamond distance is useful for several reasons. Firstly, allowing for larger entangled inputs does not change the value of the diamond distance, hence it is stable. Secondly, it has an operational meaning as determining the optimal success probability for distinguishing two unknown quantum channels $\mathcal{E}_1$ and $\mathcal{E}_2$ [32]. Equivalently, the diamond distance gives the worst-case error rate between the pair of channels. Although we will not be able to measure the diamond distance directly, we will be able to bound it in terms of measurable quantities obtainable via randomized benchmarking.

To obtain upper and lower bounds on the diamond norm, we will use the following two lemmas to relate the average gate fidelity to the trace norm of the corresponding Choi matrix and then to the diamond norm. Recall that the Choi matrix of a linear map $\Delta$ is given by $J(\Delta) = \Delta \otimes I_d(\Phi)$, where $\Phi = \sum_{ji} \ket{ij}/\sqrt{d}$ is the maximally entangled state. The first of these lemmas was proven in [20, 33].

**Lemma 6.** ([20, 33]) The average fidelity of a CPTP map $\mathcal{E}$ is related to its Choi matrix $J(\mathcal{E})$ by

$$(d + 1) F_{\text{avg}}(\mathcal{E}) = d F[\Phi, J(\mathcal{E})] + 1. \tag{31}$$

**Lemma 7.** Let $\Delta$ be a hermiticity-preserving linear map between $d$-dimensional operators. Then the following inequalities bound the diamond norm and are saturated

$$\| J(\Delta) \|_1 \leq \| \Delta \|_o \leq d \| J(\Delta) \|_1. \tag{32}$$

**Proof.** We first prove the lower bound and show that it is saturated. We have

$$\| \Delta \|_o = \sup_w \| \Delta \otimes I_d(w) \|_1 \geq \| \Delta \otimes I_d(\Phi) \|_1 = \| J(\Delta) \|_1. \tag{33}$$

To see that the above inequality is saturated, simply let $\Delta = I_d$. To prove the upper bound, we write

$$\| \Delta \|_o = \sup \left\{ \| J(\Delta) \|_1 \otimes \| \sqrt{\rho_0} \|_1 : \rho_0, \rho_1 \in D_d \right\} \tag{34}$$

which follows from theorem 6 of [34]. Using [35 theorem 9.4.5] and the triangle inequality, then the submultiplicativity of the norm, we find that

$$\| \Delta \|_o \leq \sup \left\{ \frac{1}{2} \| J(\Delta) \|_1 \otimes \| \rho_0 \|_1 + \frac{1}{2} \| J(\Delta) \|_1 \otimes \| \rho_1 \|_1 : \rho_0, \rho_1 \in D_d \right\} \leq \frac{1}{2} d \| J(\Delta) \|_1 \sup \left\{ \| \rho_0 \|_1 + \| \rho_1 \|_1 : \rho_0, \rho_1 \in D_d \right\} = d \| J(\Delta) \|_1. \tag{35}$$

To see that this bound is saturated, let $\Delta$ be the projector onto $|0\rangle \langle 0|$. \hfill $\square$
We note that it would be interesting to see if the previous bounds are still saturated when restricting the input \( \Delta \) to be a difference of channels.

6. Time-dependent gate-independent errors in randomized benchmarking

We consider the ideal case in which the noise depends only upon the time step. For such types of noise, we derive expressions for the mean \( \bar{F}_m \) and variance \( \sigma^2_m \) of the randomized benchmarking distribution \( \{F_{m,k}\} \) for fixed \( m \). In particular, we will show that for unital but nonunitary noise, \( \sigma^2_m \) decreases exponentially with \( m \), while for non-unital noise, \( \sigma^2_m \) converges to a constant dependent on the strength of the non-unitality. We will also upper-bound the variance for small \( m \), which enables the derivation of rigorous confidence intervals for the estimate of the average gate infidelity in section 4.1. We will also show that our results are stable under gate-dependent perturbations in the noise in section 8.

In order to present our results in as clear a form as possible, we will only explicitly consider the original proposal for randomized benchmarking. Interleaved benchmarking can also be treated in an almost identical manner, except that the noise is conjugated by the interleaved gate and is redefined to absorb the noise term for the interleaved gate.

Denoting the noise at time step \( t \) by \( \Lambda_t \), the sequence of operations applied to the system in the randomized benchmarking experiment with sequence \( s \in \{0,1\}^m \) is

\[
S_s = \prod_{t=m}^{0} g_t \Lambda_t.
\]

Here \( g_t \) are the ideal unitary gates which are sampled from any unitary two-design \( G \).

To make it easier to analyze the above expression, we define \( h_t = \prod_{s=m}^{t} g_s \), so that \( h_0 = 1 \), \( h_m = g_m \) and \( g_t = h_{t+1}^\dagger h_t \) for all \( t \in (0, m) \). Uniformly sampling the \( g_t \) is equivalent to uniformly sampling the \( h_t \) since \( G \) is a group; the exception is \( h_0 \) and \( g_0 \), which are chosen so that the product of all the gates is the identity (see section 2). We can then rewrite equation (36) as

\[
S_s = h_m \Lambda_m \ldots h_2^\dagger h_1 \Lambda_1 h_1^\dagger \Lambda_0 = \prod_{t=m}^{1} \Lambda_t h_t,
\]

where we incorporate the first noise term into the preparation by setting \( \rho \leftarrow \Lambda_0 \rho \). This redefinition of \( \rho \) is independent of the sequence length because we write the noise as pre- rather than post-multiplying the target. (Note that if the noise post-multiplied the target, then incorporating the final noise term in \( E \) would make \( E \) depend on the sequence length \( m \).)

The probability of observing the outcome \( E \) for the sequence \( S_s \) is \( F_{m,k} = \langle E|S_s|\rho \rangle \). We regard the set \( \{F_{m,k}\} \) as the realizations of a random variable with mean \( \bar{F}_m \) and variance \( \sigma^2_m \). Randomized benchmarking then corresponds to randomly sampling from the distribution \( \{F_{m,s}\} \) (which we henceforth refer to as the randomized benchmarking distribution) to approximate the mean \( \bar{F}_m \).

6.1. Mean of the benchmarking distribution

We now derive an expression for \( \bar{F}_m \) for general CPTP maps with time-dependent noise. A similar expression was derived for time-independent noise in [9]. We will then show how \( \bar{F}_m \)
can be used to approximate quantities of experimental interest, namely, the SPAM error, average time-dependent gate fidelity and the worst-case error due to the noise.

**Theorem 8.** The mean of the distribution \( \{ F_{m,s} \} \) for fixed \( m \) is

\[
\tilde{F}_m = E_0 \rho_0 + \tilde{E} \cdot \bar{\rho} \prod_{i=1}^{m} f_i.
\]

**Proof.** By definition, the mean is

\[
\tilde{F}_m = |q|^{-m} \sum_{s \in \mathbb{N}_q} (E \mid S_j \rangle \rangle \langle \langle S_j \mid \rho) = \left( E \mid \prod_{i=m}^{1} \Lambda^\alpha \rangle \rangle \langle \langle \rho \right),
\]

Using proposition 1 gives

\[
\tilde{F}_m = (E_0 \tilde{E}) \left( \begin{array}{c} 1 \\ 0 \\ \prod_{i=1}^{m} f_i 1_{d-1} \end{array} \right) \left( \begin{array}{c} \rho_0 \\ \bar{\rho} \end{array} \right).
\]

The parameters \( E_0 \rho_0 \) and \( \tilde{E} \bar{\rho} \) directly characterize the quality of the state and measurement procedure (with the caveat that \( \rho \) has been redefined to include a noise term), since \( \text{Tr} \rho = E_0 \rho_0 + \tilde{E} \bar{\rho} \). This can be viewed as an instance of gate set tomography using a limited number of combinations of gates [36].

The parameters \( f_i \) that give the mean of a randomized benchmarking distribution are closely related to an operational characterization of the amount of noise, namely, the average gate infidelity [10, 12] (which gives the average error rate), as

\[
r_i = 1 - F_{\text{avg}}(\Lambda_i) = \frac{d-1}{d} \left( 1 - f_i \right).
\]

The randomized benchmarking protocol will enable the estimation of \( \prod_{i} f_i \), which can then be used to estimate the average gate infidelity averaged over arbitrary time intervals (as shown in section 4.2).

We now show that the average gate infidelity provides an upper and a lower bound on \( \frac{1}{2} \| \Lambda - 1 \|_6 \), which gives the worst-case error introduced by using \( \Lambda \) instead of \( 1 \). An upper bound of the same form was stated without proof in [22], however, the bound here is a factor of two better. The following relation between the diamond distance and the average gate fidelity can also be applied at each time step to relate the time-averaged average gate fidelity to the average diamond distance from the identity channel. Note that the following bound is very loose in the regime \( m r \ll 1 \) (since in that regime, \( r \ll \sqrt{r} \)), which is also the regime in which we will typically use it.

**Proposition 9.** Let \( r = 1 - F_{\text{avg}}(\Lambda) \) be the average error rate for \( \Lambda \). Then

\[
r(d+1)/d \leq \frac{1}{2} \| \Lambda - 1 \|_6 \leq \sqrt{d(d+1)r}.
\]
Proof. Applying lemma 7 to $\Delta = \Lambda - 1$ gives
\[
D[\Phi, J(\Lambda)] = \frac{1}{2} \| J(\Lambda) - \Phi \| \lesssim \frac{1}{2} \| \Lambda - 1 \| \lesssim \frac{d}{2} \| J(\Lambda) - \Phi \| = d D[\Phi, J(\Lambda)].
\] (43)
Recalling that $\Phi$, the maximally entangled state, is a pure state and using equation (26) and (27) gives
\[
1 - F[\Phi, J(\Lambda)] \leq D[\Phi, J(\Lambda)] \lesssim \sqrt{1 - F[\Phi, J(\Lambda)]}.
\] (44)
From lemma 6, $1 - F[\Phi, J(\Lambda)] = d^{-1}(d + 1)r$. Substituting this into the above expression and combining the inequalities completes the proof.

6.2. Upper bounds on the variance

We now consider the variance $\sigma_m^2$ of the distribution $\{F_{m,k}\}$ for fixed $m$. It has been observed that the standard error of the mean (and hence the sample variance) can be remarkably small in experimental applications of randomized benchmarking using relatively few random sequences [12, 22]. In this section, we will prove that the variance due to sampling random sequences is indeed small in scenarios of practical interest (i.e., $mr \ll 1$) by obtaining an upper bound on $\sigma_m^2$ in terms of $mr$. For the special case of a qubit, we will also obtain a significantly improved upper bound in terms of $m$ and $r$.

We begin by obtaining a general bound on $\sigma_m^2$ that depends only on $m$, $r$ and the dimension $d$ of the system being benchmarked. In order to present results in a simple form, we assume that the noise is time- and gate-independent, however, the results in this section can readily be generalized to time-dependent noise.

As a first attempt at obtaining a good bound on the variance, we use only the fact that when $mr \ll 1$, we have $\bar{F}_m \approx A + B$, where $A = E_0\rho_0$ and $B = \bar{E} \cdot \bar{\rho}$. Expanding the expression from theorem 8 to first order in $r$ using equation (41) gives
\[
\bar{F}_m = A + B - \frac{Bmdr}{d - 1}.
\] (45)
The value of all realizations of $\bar{F}_m$ (i.e., the probabilities $F_{m,s}$) are all in the unit interval. Since the distribution with the largest variance that has mean $\bar{F}_m$ and takes values in the unit interval is the binomial distribution with that mean, we then have
\[
\sigma_m^2 \leq \bar{F}_m (1 - \bar{F}_m) = (A + B)(1 - A - B) + \frac{mdBr}{d - 1}.
\] (46)
While simple to obtain, this bound has a constant off-set term that depends upon the SPAM which seems to be unavoidable. This term would be zero in the absence of SPAM, and could even be eliminated if the probabilities $F_{m,s}$ could be restricted to the interval $[1 - A - B, A + B]$. However, as illustrated in section 4.1, this cannot be done in general. Moreover, we expect that the above argument substantially overestimates the variance because it ignores the possibility that many sequences may have $F_{m,s}$ closer to $\bar{F}_m$.

We now obtain an alternative bound that has a larger coefficient for $r$, but no constant term. We note from the outset that the following bound is not tight in general (and the previous bound suggests that the dimensional factor is an artifact of the proof technique), though by
improving one of the steps we will be able to obtain a tight bound for qubits. To facilitate our analysis, we use the identity $(E \otimes \mathcal{I}) \rho = (E \otimes \mathcal{I}) \otimes (\mathcal{I} \otimes \rho) = (E \otimes \mathcal{I}) \otimes (\mathcal{I} \otimes \rho) = (E \otimes \mathcal{I}) \rho$ to write the variance as

$$\sigma_m^2 = \left| G^{-m} \sum_k F_{m,k}^2 - \bar{F}_m^2 \right| = \left( E \otimes \mathcal{I} \right) \left[ \left( \left( A \otimes \mathcal{I} \right) \rho \right)^m - \left( \left( A \otimes \mathcal{I} \right) \rho \right)^m \right] \rho.
$$  

(47)

**Theorem 10.** The variance for time- and gate-independent randomized benchmarking of $d$-level systems with time- and gate-independent noise satisfies

$$\sigma_m^2 \leq 4d(d + 1)mr + O(m^2r^2d^4).$$  

(48)

**Proof.** We write $\Lambda = \mathbb{I} - r\Delta$, where the first row of $\Delta$ is zero since $\Lambda$ is CPTP. Since $(d^2 - 1)\Delta = \text{Tr} \varphi = \text{Tr} \Lambda - 1$, we can use equation (41) to obtain $\text{Tr} \Delta = d(d + 1)$ We then expand the expression

$$\sigma_m^2 = \left( E \otimes \mathcal{I} \right) \left[ \left( \left( A \otimes \mathcal{I} \right) \rho \right)^m - \left( \left( A \otimes \mathcal{I} \right) \rho \right)^m \right] \rho.
$$  

(49)

to second order in $r\Delta$. Note that $(\Delta \otimes \mathbb{I}) = \Delta^g \otimes \mathbb{I}$ and so all the first-order terms and the second-order terms where the $\Delta$ act at different times will cancel. Therefore the only second-order terms are the $m$ terms with $\Delta^g \otimes \mathbb{I}$ and so the variance is

$$\sigma_m^2 = mr^2 \left( E \otimes \mathcal{I} \right) \left[ \left( \left( A \otimes \mathcal{I} \right) \rho \right)^m - \left( \left( A \otimes \mathcal{I} \right) \rho \right)^m \right] \rho + O(r^3\Delta^3).
$$  

(50)

Noting that $\Delta^g = \frac{d(d+1)}{d^2-1} \mathbb{I}$, the variance satisfies

$$\sigma_m^2 \leq mr^2 \left| G^{-1} \right| \left| \sum_{g \in G} \left( \Delta^{g^2} \right)^m \right| \rho + O(r^3\Delta^3) + O(mr^2)
$$  

$$\leq mr^2 \left| \Delta^{g^2} \right| \rho + O(r^3\Delta^3) + O(mr^2)
$$  

$$\leq mr^2 \left| \Delta \right| \rho + O(r^3\Delta^3)
$$  

$$\leq 4d(d + 1)mr + O(r^3\Delta^3) + O(mr^2),
$$  

(51)

where we have used the triangle inequality, the invariance of the diamond norm under unitary conjugation, the submultiplicativity of the diamond norm [with $\Delta \otimes \Delta = (\Delta \otimes \mathbb{I})(\mathbb{I} \otimes \Delta)$] and proposition 9.

Finally, consider terms of $O(r^k\Delta^k)$ for $k > 2$. For $k \geq 3$, all $O(m^k)$ such terms are upper-bounded by $r^k \left| \Delta^k \right| \rho$ and so are $O(r^{k/2}d^k)$. Therefore the only contributions of $O(r^{k^2})$ or greater are from $k = 3$ and $k = 4$.

For $k = 3$, the only terms that will not cancel are products whose only nontrivial terms are a $(\Delta \otimes \Delta)^g$ and a $\Delta^g \otimes \mathbb{I} = \frac{d(d+1)}{d^2-1} \mathbb{I}$. There are only $O(m^2)$ such terms, and applying the diamond norm bound to $\frac{d}{d^2-1} (\Delta \otimes \Delta)^g$ shows that such terms contribute at most $O(m^2r^2d^2)$.

For $k = 4$, the only terms that will be of $O(r^2)$ are those that are products with two $(\Delta \otimes \Delta)^g$ terms. Again, there are only $O(m^2)$ such terms and so such terms also contribute at most $O(m^2r^2d^4)$. $\square$
While the bound in theorem 10 is promising, it is not sufficiently small to justify the sequence lengths chosen in many experimental implementations of randomized benchmarking for a single qubit, since \( mr \approx 10^{-2} \) in many such experiments and so the contribution to standard error of the mean due to sampling random gate sequences is expected to be on the order of \( 0.1 \frac{K^{-1/2}}{m} \), where \( K \) is the number of random sequences of length \( m \) that are sampled.

One of the loosest approximations in theorem 10 is the use of the triangle inequality to upper-bound the contribution from terms of the form \( |\Delta \otimes g|^2 \). Avoiding this is difficult in general, however, for the case of a single qubit, we can significantly improve the following bound by understanding the irrep structure of the representation \( g \otimes g \). This irrep structure will depend on the choice of two-design, so we now fix the two-design to be the single qubit Clifford group, \( C_2 \) and work in the Pauli basis \( \mathcal{A} = \{I, X, Y, Z\}/\sqrt{2} \) (where the factor of \( \sqrt{2} \) makes the basis trace-orthonormal). In particular, we will work in the block basis

\[
\left( \begin{array}{c}
1 \\
1 \otimes \vec{\sigma} \\
\vec{\sigma} \otimes 1 \\
\vec{\sigma} \otimes \vec{\sigma}
\end{array} \right),
\]

where \( \vec{\sigma} = \{X, Y, Z\}/\sqrt{2} \). Restricting the Liouville representation to each of the blocks in the above basis will give a rep of \( C_2 \), where the first three reps have already been characterized. We now characterize the final subrep, \( (\phi \otimes 2, C^9) \).

**Proposition 11.** The representation \( (\phi \otimes 2, C^9) \) of \( C_2 \) is the direct sum of four inequivalent irreps.

**Proof.** The proof follows from a direct application of Schur’s orthogonality relations, which imply

\[
|C_2|^{-1} \sum_{g \in C_2} \chi_{\phi \otimes 2}^*(g) \chi_{\phi \otimes 2}(g) = \sum_{\lambda} n_{\lambda}^2,
\]

where \( n_{\lambda} \) is the multiplicity of the irrep \( \lambda \) in the rep \( \phi \otimes 2 \).

The character is given by

\[
\chi_{\phi \otimes 2}(g) = \text{Tr} \ g \otimes 2 = (\text{Tr} \ g)^2.
\]

Since the elements of \( C_2 \) permute Paulis (up to signs), the diagonal elements of \( G \) in the Pauli basis are either 1 or \(-1\) and there are 0, 1 or 3 diagonal elements that can contribute to \( \text{Tr} \ g \).

There are eight elements of \( C_2 \) with no diagonal elements, namely, the eight permutations \( X \rightarrow \pm Y \rightarrow \pm Z \) and \( X \rightarrow \pm Z \rightarrow \pm Z \). There is one element with all diagonal elements equal, namely, the identity (note that \(-1\) is antiunitary so is not in the Clifford group). All other 15 elements of the Clifford group have \( \chi_{\phi}(g) = \pm 1 \) since the diagonal elements cannot sum to any values in \( \{0, \pm 2, \pm 3\} \).

Plugging these character values into equation (53) gives

\[
|C_2|^{-1} \sum_{g \in C_2} \chi_{\phi \otimes 2}^*(g) \chi_{\phi \otimes 2}(g) = \frac{1}{24} \sum_{g \in C_2} \left| \chi_{\phi}(g) \right|^4 = \frac{1}{24} (3^4 + 15) = 4.
\]

Given that the multiplicity of an irrep must be a nonnegative integer, there are two possibilities. Either there are four inequivalent irreps or the rep \( \phi \otimes 2 \) contains two equivalent irreps. By
proposition 2, $\varphi \otimes 2$ contains a trivial irrep with multiplicity 1 and so cannot contain two equivalent irreps, therefore there are 4 inequivalent irreps.

The following bases of operators

$$A_1 = \frac{1}{2\sqrt{3}} (XX + YY + ZZ),$$

$$A_2 = \left\{ \frac{1}{2\sqrt{2}} (XX - YY), \frac{1}{2\sqrt{6}} (XX + YY - 2ZZ) \right\},$$

$$A_5 = \frac{1}{2\sqrt{2}} \left\{ XY - YX, XZ - ZX, YZ - ZY \right\},$$

$$A_T = \frac{1}{2\sqrt{2}} \left\{ XY + YX, XZ + ZX, YZ + ZY \right\},$$

span the four irreps. □

The fact that $g \otimes g$ is a direct sum of four inequivalent irreps will allow us to use Schur’s lemma on the unital block of $\Delta$. To account for the nonunital component, we use the following bound.

**Proposition 12.** For any CPTP qubit channel $\Lambda: D_2 \to D_2$ with average gate infidelity $r < 1/3$, the nonunital part $\alpha$ obeys the inequality

$$\| \alpha \|_2^2 \leq 9r^2. \quad (57)$$

**Proof.** Any trace-preserving qubit channel as

$$\Lambda = (1 \oplus U) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & w_1 & 0 & 0 \\ 0 & 0 & w_2 & 0 \\ t & 0 & 0 & w_3 \end{pmatrix} (1 \oplus U^\dagger)(1 \oplus V). \quad (58)$$

for some $U, V \in O(3)$ [corresponding to unitaries $u, v \in U(2)$], where $|w_j|$ are the singular values of $\varphi$ with $|w_j| \in [0, 1]$ for all $j$ [24] and we have added the $(1 \oplus U^\dagger)$ term for convenience. By Von Neumann’s trace inequality [37]

$$3 - 6r = \text{Tr} \varphi = |\text{Tr} \ UWU^\dagger V| \leq \sum_j |w_j|, \quad (59)$$

where $W = \text{diag}(w_1, w_2, w_3)$ and we have used the fact that the singular values of $V$ are all one. For notational convenience, we define perturbations $\delta_j$ by $|w_j| = 1 - \delta_j r$ which then satisfy $\sum_j \delta_j \leq 6$ and $\delta_j \geq 0$ for all $j$.

The conditions for $\Lambda$ to be completely positive are

$$|t| + |w_3| \leq 1$$

$$\left( w_j \pm w_k \right)^2 \leq \left( 1 \pm w_l \right)^2 \quad (60)$$
for any permutation \( \{j, k, l\} \) of \( \{1, 2, 3\} \). Therefore
\[
|\ell| \leq 1 - |w_3| = \delta_3 r,
\]
and, since \( \| \alpha \|_2 \) is invariant under the unitary transformations in equation (58), \( \| \alpha \|_2 = |\ell| \). Therefore the only remaining problem is to bound \( \delta_3 \) (note that at this point, we could accept the trivial bound \( \delta_3 \leq 6 \)).

If \( w_3 < 0 \), then complete positivity implies
\[
(2 - \delta_1 r - \delta_2 r)^2 \leq \delta_3^2 r^2,
\]
which cannot be satisfied subject to \( \sum_j \delta_j \leq 6 \) and \( \delta_j \geq 0 \) for \( r < 1/3 \). Therefore for all \( r < 1/3 \), \( w_3 = 1 - \delta_3 r > 0 \).

Considering the conditions
\[
(\delta_j - \delta_k)^2 \leq \delta_1^2
\]
for all permutations \( \{j, k, l\} \) of \( \{1, 2, 3\} \), we see that \( \delta_3 \leq \max_j \delta_j \leq 3 \) and so \( \| \alpha \|_2 \leq 3r \). □

Combining the irrep structure of \( g \otimes g \) and the bound on the nonunital component allows us to improve the bound in theorem 10 for the special case of one qubit. As discussed in section 4, the following bound provides a rigorous justification of current experiments and allows values of \( K_m \) to be chosen that are substantially smaller than previously justified rigorously, that is, \( K_m \approx 145 \) as opposed to \( K_m \approx 7 \times 10^4 \) as estimated in [10].

**Theorem 13.** The variance for arbitrary time- and gate-independent noise satisfies
\[
\sigma_m^2 \leq m^2 r^2 + \frac{7}{4} m^2 r^2 + 6\delta m r + O(m^2 r^3) + O(\delta m^2 r^2),
\]
where \( \delta = |\delta_E \cdot \delta_\rho| \leq 1/2 \) for any choice of \( \vec{w} \in \{\vec{x}, \vec{y}, \vec{z}\} \) and \( \delta_\rho, \delta_E \perp \vec{w} \) such that
\[
\vec{E}^T = a\vec{w} + \delta_E
\]
\[
\vec{\rho} = b\vec{w} + \delta_\rho.
\]

**Proof.** To prove the theorem, we will derive an exact expression for the variance and then approximate it in the relevant regimes.

We begin by noting that in the basis \( \{1^\otimes 2, 1 \otimes A, A \otimes 1, A \otimes A\} \) we have
\[
(A^\otimes 2)^{C_2} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & \phi^{C_2} & 0 & 0 \\
0 & 0 & \phi^{C_2} & 0 \\
\end{pmatrix},
\]
where \( P_1 = \sum_{g \in C_2} g^{\otimes 2} \). It can be verified that \( E^\otimes 2 \) is in the null space of \( \sum_{g \in C_2} g^{\otimes 2} \otimes g \alpha \) and \( \sum_{g \in C_2} g \alpha \otimes \phi^{(g)} \) for any \( E \) by, for example, considering a basis for the space of \( \phi^{(g)} \)’s. We
note in passing that this property is not a general property of two-designs, in that it does not hold for the single-qutrit Clifford group.

By propositions 11 and \( \lfloor \omega \rfloor \), \( \sum_{\lambda}^{R} \lambda_{R} P_{R} \) where the \( P_{R} \) are the projectors onto the irreps from proposition 11 and \( \lambda_{R} = \text{Tr} P_{R} \omega \otimes 2 / \text{Tr} P_{R} \). From equation (47), together with the orthogonality of the projectors \( P_{R} \), we have

\[
\sigma_{m}^{2} = P_{1}^{2} E \otimes 2 \sum_{i=0}^{m-1} \lambda_{i}^{f} + \sum_{R}^{m} \lambda_{R}^{m} E \otimes 2 P_{R} \rho \otimes 2 - (1 - 2r)^{2m}(E \rho)^{2}. \tag{67}
\]

We can bound the first term using

\[
\rho_{1}^{2} E \otimes 2 P_{1} \sum_{i=0}^{m-1} \lambda_{i}^{f} \leq \frac{1}{3} \rho_{0}^{2} \| E \|^{2} \| \rho \|^{2} \sum_{i=0}^{m-1} 1 \leq \frac{3mr^{2}}{4}, \tag{68}
\]

where we have used and the trivial bound \( \lambda_{i} \leq 1 \) (for the first term only) and proposition 12 to obtain the final inequality.

Similarly, the eigenvalues can be calculated to be

\[
\lambda_{1} = \frac{1}{3} \sum_{j,k} q_{j,k}^{2} = \frac{1}{3} \text{Tr}(\omega \omega) \]
\[
\lambda_{2} = \frac{1}{3} \sum_{j} q_{j,j}^{2} - \frac{1}{6} \sum_{j \neq k} q_{j,k}^{2} = \frac{1}{6} \sum_{j} q_{j,j}^{2} - \frac{1}{6} \text{Tr} \omega \omega \]
\[
\lambda_{T} = \frac{1}{6} \sum_{j \neq k} (q_{j,j} + q_{j,k}) = \frac{1}{6} \text{Tr}(\omega^{2}) + \frac{1}{6} (\text{Tr} \omega)^{2} - \frac{1}{3} \sum_{j} q_{j,j}, \tag{69}
\]

where we have omitted \( \lambda_{S} \) since it will not contribute to the variance since any symmetric vector (such as \( E \otimes 2 \)) will be orthogonal to \( P_{S} \).

We now consider general noise with \( E \) and \( \rho \) as in equation (65), where, without loss of generality, we set \( \omega = z \). We begin by considering the case \( \delta = 0 \), for which \( E \otimes 2 P_{1} = E \otimes 2 P_{2} = 0 \). Then a simple calculation using proposition 12 gives \( E \otimes 2 P_{1} \rho \otimes 2 = a^{2} b^{2}/3, (E \rho)^{2} = a^{2} b^{2} \) and \( E \otimes 2 P_{2} \rho \otimes 2 = \frac{2}{3} a^{2} b^{2} \).

The eigenvalues \( \lambda_{1} \) and \( \lambda_{2} \) can be written as \( x + 2y \) and \( x - y \) respectively, where \( x = \frac{1}{3} \sum_{j} q_{j,j} \) and \( y = \frac{1}{6} \sum_{j \neq k} q_{j,k}^{2} \). Writing \( \varphi = 1 - \Delta r \), where \( \text{Tr} \Delta = 6 \) (see the discussion in the proof of theorem 10), we have

\[
1 - 4r + 4r^{2} \leq x := \frac{1}{3} \sum_{j} q_{j,j}^{2} = 1 - 4r + \frac{r^{2}}{3} \sum_{j} \Delta_{j}^{2} \leq 1 - 4r + 12r^{2}, \tag{70}
\]

where the maximum and the minimum are obtained by maximizing and minimizing \( \sum_{j} \Delta_{j}^{2} \) subject to \( \sum \Delta_{j} = 6 \) for real matrices \( \Delta \) with nonnegative diagonal entries respectively. The diagonal entries of \( \Delta \) must be nonnegative since all entries of \( \varphi \) have modulus upper-bounded by 1 (which can be easily verified from the form of extremal channels in equation (58)). Therefore the variance satisfies
\[ \sigma_m^2 \leq \frac{3mr^2}{4} + \frac{a^2b^2}{3} [(x + 2y)^m + 2(x - y)^m - 3(1 - 2r)^{2m}] . \]  

(71)

Since \( 1 \geq \lambda_1 - 2y = x \geq 1 - 4r \) by equation (70), we have \( y \leq 2r \) and so, using a binomial expansion to \( O(r^3) \) gives

\[ (x + 2y)^m + 2(x - y)^m - 3(1 - 2r)^{2m} \leq 12mr^2 + 12m^2r^2 + O(m^3r^3). \]  

(72)

Noting that \( a^2b^2 \leq 1/4 \) gives

\[ \sigma_m^2 \leq m^2r^2 + + \frac{7}{4}mr^2 + O(m^2r^3). \]  

(73)

We now consider the correction when \( \delta > 0 \) in equation (65), which will realistically always be the case since \( \rho \) incorporates a residual noise term. Then we define functions \( h_R(\vec{\delta}_1, \vec{\delta}_2) \) by

\[ \vec{E} \otimes^2 P_R\vec{\rho} \otimes^2 = a^2b^2z^2P_Rz^2 + h_R(\vec{\delta}_1, \vec{\delta}_2), \]  

(74)

where we will henceforth omit the arguments of \( h_R \). Since \( \sum_R \vec{E} \otimes^2 P_R\vec{\rho} \otimes^2 = (\vec{E}\vec{\rho}) \), we can write the variance as

\[ \sigma_m^2 \leq a^2b^2\sigma_{m,z}^2 + \sum_R h_R \left[ \lambda_R^m - (1 - 2r)^{2m} \right]. \]  

(75)

To \( O(r^2) \), the smallest eigenvalue is \( \lambda_2 \), since

\[ \frac{1}{6}\sum_{j \neq k} q_{j,j}q_{k,k} = \frac{1}{3}\sum_j q_{j,j}^2 + O(r^2) \]

\[ \frac{1}{6}\left| \sum_{j \neq k} q_{j,k}q_{k,j} \right| \leq \frac{1}{6}\sum_{j \neq k} q_{j,k}^2, \]  

(76)

where the first line follows by writing \( q_{j,k} = 1 - r\Delta_{j,k} \) and the second from the inequality \( q_{j,k}^2 + q_{k,j}^2 \geq 2|q_{j,k}q_{k,j}| \) and the triangle inequality. Therefore, to \( O(r^2) \),

\[ 1 - 8r \leq x - y = \lambda_2 \leq \lambda_R \leq 1 \]  

for all \( R \) (where the bounds on \( x \) and \( y \) are as in equation (70)) and so \( \left| \lambda_R^m - (1 - 2r)^{2m} \right| \leq 4mr + O(m^2r^2) \) for all \( R \). Therefore

\[ \sigma_m^2 \leq a^2b^2\sigma_{m,z}^2 + [4mr + O(m^2r^2)] \sum_R h_R \]

\[ \leq m^2r^2 + \frac{7}{4}mr^2 + [4mr + O(m^2r^2)] \sum_R h_R + O(m^2r^3) \]

\[ \leq m^2r^2 + \frac{7}{4}mr^2 + 6\delta mr + O(m^2r^3) + O(\delta m^2r^2), \]  

(77)

where we have obtained the final inequality using

\[ a^2b^2 + \sum_R h_R = (\vec{E}\vec{\rho})^2 = a^2b^2 + 2ab(\vec{\delta}_1 \cdot \vec{\delta}_2) + (\vec{\delta}_1 \cdot \vec{\delta}_2)^2 \leq a^2b^2 + \frac{3\delta}{2}. \]  

(78)

where the final inequality follows since \( \delta = |\vec{\delta}_1 \cdot \vec{\delta}_2|, |ab| \leq 1/2. \)  

\[ \square \]
It is worth noting that one could in principle fill in the implicit constants given in the big-\(O\) notation by following the previous argument with sufficient care. To have a truly rigorous confidence region, one would need to take this into account, but for current parameter regimes of interest, the terms really are negligible, so it hardly seems worth optimizing this concern.

We also note that \(\delta_e\) will typically have entries of order \(\sqrt{r}\) even without SPAM, since the off-diagonal terms for generic noise are of order \(\sqrt{r}\) and there is a residual noise term that has been incorporated into \(\rho\). However, the corresponding entries in \(\delta_E\) will generally be smaller (or at least, are determined only by SPAM).

We now show that the variance can be even further improved (by a factor of \(m\) and with no dependence on the state and measurement) for noise that is diagonal in the Pauli basis.

**Corollary 14.** If the unital block of the noise is diagonal in the Pauli basis, this bound can be improved to

\[
\sigma_m^2 \leq \frac{11mr^2}{4} + O(m^2r^3). \tag{79}
\]

**Proof.** For noise such that \(\varphi\) is diagonal in the Pauli basis, \(\lambda_1 = \lambda_2 = x\) and \(\lambda_T \leq \lambda_1\), which can be shown using the inequality \(2ab \leq a^2 + b^2\) for \(a, b \in \mathbb{R}\). Therefore, for noise that is diagonal in the Pauli basis, we have

\[
\sigma_m^2 \leq \frac{3mr^2}{4} + \frac{1}{4} [(1 - 4r + 12r^2)m - (1 - 2r)^2m] \\
\leq \frac{11mr^2}{4} + O(m^2r^3) \tag{80}
\]

by equation (70).

One consequence of the above corollary is that the variance of the randomized benchmarking distribution will typically depend strongly upon the choice of two-design even for gate independent noise. This observation follows from the above theorem by noting that the unital block can be perturbed by an arbitrarily small amount to allow it to be unitarily diagonalized. Performing randomized benchmarking in the basis where the unital block is diagonalized (i.e., setting \(G = C_x^U\)) will give variances of order \(mr^2\), while randomized benchmarking in other bases will give variances of order \(m^2r^2\).

### 7. Asymptotic variance of randomized benchmarking

We now consider the variance \(\sigma_m^2\) of the distribution \(\{F_{m,s}\}\) as \(m \to \infty\). While not directly relevant to current experiments, the asymptotic behavior is nevertheless interesting in that it may provide a method of estimating the amount of nonunitality.

We will prove that, for the class of channels defined below called \(n\)-contractive channels (which are generic in the space of CPTP channels), \(\sigma_m^2\) decays exponentially in \(m\) to a constant that quantifies the amount of nonunitality. Unfortunately, we will not be able to provide a bound on the decay rate. In fact, no such bound is possible without further assumptions since the channel \([(1 - \epsilon)U + \epsilon E]^G\) for any unitary \(U\) and two-contractive channel \(E\) will have an eigenvalue \(1 - \epsilon + O(\epsilon) < 1\) corresponding to the trivial subrep (this can be seen by following...
the proof of proposition 16). This eigenvalue will result in a variance that decays as $(1 - \epsilon)^m$ for arbitrary $\epsilon > 0$.

**Definition 15.** A channel $\Lambda: D_d \to D_d$ is $n$-contractive with respect to a group $G \subseteq U(d)$ if $(\Lambda^{\otimes n})^G$ has at most one eigenvalue of modulus 1.

We now prove that all unital but nonunitary channels are two-contractive with respect to any finite two-design. We conjecture that all nonunitary channels are in fact two-contractive with respect to any unitary two-design. An equivalent statement for trace-preserving channels $\Lambda$ is that $(\Lambda^{\otimes 2})^G$ is strongly irreducible whenever $\Lambda$ is not unitary [38]. As a corollary of the following proposition, this conjecture holds for qubits, since, for qubits, the projection onto the unital part of a CPTP map is also a CPTP map [21]. However, proving it for higher dimensions remains an open problem.

**Proposition 16.** Let $\Lambda$ be a completely positive, trace-preserving and unital channel and $G$ a unitary two-design. Then $\Lambda$ is two-contractive with respect to $G$ if and only if it is nonunitary.

**Proof.** First assume $\Lambda$ is unitary. Since $(\varphi, \mathbb{R}^{d^2-1})$ is an orthogonal irrep of $U(d)$, $(\varphi, \mathbb{R}^{d^2-1})^G$ contains the trivial rep as a subrep with multiplicity 1 by proposition 2. Therefore for any $U \in U(d)$ and in a fixed Schur basis (i.e., independent of $U$), $\varphi(U)^{\otimes 2} = 1 \oplus \mathcal{T}(U)$ for some homomorphism $\mathcal{T}$. Therefore any vector $v$ in the (one-dimensional) trivial representation is a +1-eigenvector of $\varphi^{\otimes 2}$ for any $U$ and consequently is a +1-eigenvector of $[(\varphi^{\otimes 2}(\Lambda))^G]$. We now show that for all completely positive, trace-preserving and unital $\Lambda$

$$
\left\| (\varphi^{\otimes 2})^G \right\|_\infty^2 \leq 1 - |G|^{-1} \left[ 1 - \frac{\text{Tr} \varphi^{\dagger} \varphi}{d^2 - 1} \right].
$$

(81)

Recall that one of the equivalent definitions of the spectral norm is

$$
\left\| (\varphi^{\otimes 2})^G \right\|_\infty^2 = \max_{u: \|u\|_1 = 1} \left( (\varphi^{\otimes 2})^G \right)^\dagger \left( (\varphi^{\otimes 2})^G \right) u.
$$

(82)

Expanding the averages over $G$ gives

$$
\left\| (\varphi^{\otimes 2})^G \right\|_\infty^2 = \max_{u: \|u\|_1 = 1} \left| G \right|^{-2} \sum_{g,h \in G} u^\dagger \left[ \left( (\varphi^{\otimes 2})^g \right)^\dagger \left( (\varphi^{\otimes 2})^h \right) u \right]
$$

$$
\leq 1 - |G|^{-1} + \max_{u: \|u\|_1 = 1} |G|^{-2} \sum_{g \in G} u^\dagger \left[ \left( (\varphi^{\dagger} \varphi)^{\otimes 2} \right)^g u \right],
$$

(83)

where in the second line we have used the improved bound for unital channels in proposition 5 to bound the contribution from the $|G|^2 - |G|$ terms with $g \neq h$.

Now let $u$ be an arbitrary unit vector and write $u = \sum_{j,k} u_{j,k} v_j \otimes v_k$, where $\{v_j\}$ is an orthonormal basis of $\mathbb{C}^d$. Then, since $\left( (\varphi^{\dagger} \varphi)^g \right)^{\otimes 2}$ is positive semidefinite with eigenvalues upper-bounded by 1 by proposition 5, we have
\[
\begin{align*}
\sum_{j,k} \left| u_{j,k} \right|^2 \left[ v_j \left( \varphi \right)^g \right] v_j &\leq \sum_{j,k} \left| u_{j,k} \right|^2 \left[ v_k \left( \varphi \right)^g \right] v_j, \\
&= \sum_{j,k} \left| u_{j,k} \right|^2 \left[ v_j \left( \varphi \right)^g \right] v_j, \quad (84)
\end{align*}
\]

where we have used \( 0 \leq v_k \left( \varphi \right)^g v_k \leq 1 \) for all \( k \) and \( g \) to obtain the second line. By proposition 1

\[
\left| \mathcal{G} \right|^{-1} \sum_{g \in \mathcal{G}} \left( \varphi \right)^g = \frac{\text{Tr} \left( \varphi \right)}{d^2 - 1}, \quad (85)
\]

so

\[
\left| \mathcal{G} \right|^{-1} \sum_{g \in \mathcal{G}} \left( u \right)^g \left( \varphi \right)^g u \leq \frac{\text{Tr} \left( \varphi \right)}{d^2 - 1} \sum_{j,k} \left| u_{j,k} \right|^2 \leq \frac{\text{Tr} \left( \varphi \right)}{d^2 - 1}, \quad (86)
\]

for all \( u \) such that \( \| u \|_2 = 1 \). Therefore

\[
\left\| \left( \varphi^{\otimes 2} \right)^G \right\|_\infty^2 \leq 1 - \left| \mathcal{G} \right|^{-1} \left[ 1 - \frac{\text{Tr} \left( \varphi \right)}{d^2 - 1} \right]. \quad (87)
\]

Theorem 17. Let \( \Lambda \) be a two-contractive channel with respect to a group \( \mathcal{G} \) that is also a two-design. Then the variance due to sampling random gate sequences of elements from \( \mathcal{G} \) decays exponentially to

\[
\frac{\rho_0^2 \mathcal{E}^{\otimes 2} P_i \alpha^{\otimes 2}}{1 - \lambda_1}, \quad (88)
\]

where \( P_i = \left| \mathcal{G} \right|^{-1} \sum_{g \in \mathcal{G}} g^{\otimes 2} \) is a rank-1 projector and \( \lambda_1 = \text{Tr} P_i \left( \varphi \right)^{\otimes 2} \).

Proof. For convenience, we use the block basis \( \{ 1^{\otimes 2}, 1 \otimes A, A \otimes 1, A \otimes A \} \) for the matrix representation. In this basis, we can write
where we have used \( \sum_{g \in \mathcal{G}} g^8 = 0 \) by proposition 1, \( P_1 = |\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} g^{\otimes 2} \), and

\[
\begin{align*}
\begin{bmatrix} \sum \phi_{(g)} & \mathcal{G} \end{bmatrix} &= \begin{bmatrix} \sum \phi_{(g)} & \mathcal{G} \end{bmatrix} \mathcal{G}^{-1} \sum_{g \in \mathcal{G}} g^{\otimes 2} \\
\end{align*}
\]

By propositions 1 and 2, \( P_1 \) is a rank-1 projector onto the trivial subrep, which occurs with multiplicity 1.

It can easily be shown using an inductive step that

\[
\begin{bmatrix} (A^{\otimes 2})^m \end{bmatrix} = \begin{bmatrix} (A^{\otimes 2})^m \end{bmatrix} = \begin{bmatrix} \begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & \Lambda & 0 & 0 \\
0 & 0 & \Lambda & 0 \\
0 & 0 & 0 & \Lambda \\
\end{array} \end{bmatrix}
\]

where

\[
\begin{align*}
A_m &= \sum_{i=0}^{m-1} (\phi^{\otimes 2}) P_1 \alpha^{\otimes 2} \\
B_m &= \sum_{i=0}^{m-1} f^{m-1-i} \left[ (\phi^{\otimes 2})^i \right] b \\
C_m &= \sum_{i=0}^{m-1} f^{m-1-i} \left[ (\phi^{\otimes 2})^i \right] c.
\end{align*}
\]

Since the trivial subrep occurs with multiplicity 1 and \( (\phi^{\otimes 2}) \mathcal{G} \) commutes with \( g^{\otimes 2} \) for all \( g \), by proposition 1 we can write \( (\phi^{\otimes 2}) \mathcal{G} = \lambda_1 P_1 + M \) for some matrix \( M \) orthogonal to \( P_1 \), where \( \lambda_1 = \text{Tr} P_1 (\phi^{\otimes 2}) \mathcal{G} \). Therefore we have

\[
A_m = P_1 \alpha^{\otimes 2} \sum_{i=0}^{m-1} \lambda_1^i.
\]
Substituting these expressions into equation (47) gives
\[
\sigma_m^2 = \rho_0^2 E^\otimes 2 P_1 \alpha^\otimes 2 \sum_{r=1}^{m} \lambda_r^i \rho + \rho_0^2 \tilde{E}^\otimes 2 \left( B_m + C_m \right) \tilde{\rho} \\
+ \tilde{E}^\otimes 2 \left[ (\varphi^\otimes 2)^C \right]^m \tilde{\rho}^\otimes 2 - f^{2m} \left( \tilde{E}^\rho \right)^2.
\] (94)

We now prove that all but the first term decay exponentially for any two-contractive channel with respect to \( \mathcal{G} \). Let \( SJS^{-1} \) be the Jordan decomposition of \((\varphi^\otimes 2)^C\). Then, by the submultiplicativity of the spectral norm and a standard identity
\[
\left\| \left( \varphi^\otimes 2 \right)^C \right\|_\infty^m = \left\| (SJS^{-1})^m \right\|_\infty \\
\leq \left\| J^m \right\|_\infty \left\| S \right\|_\infty \left\| S^{-1} \right\|_\infty \\
\leq (d^2 - 1)^2 \left\| J^m \right\|_{\text{max}} \left\| S \right\|_\infty \left\| S^{-1} \right\|_\infty,
\] (95)

where \( \left\| M \right\|_{\text{max}} = \max_{j,k} \left| M_{j,k} \right| \) and \( J \) is a \((d^2 - 1)^2 \times (d^2 - 1)^2\) matrix. Note that since \( S \) is invertible, both \( \left\| S \right\|_\infty \) and \( \left\| S^{-1} \right\|_\infty \) are finite.

By explicit calculation
\[
\left\| J^m \right\|_{\text{max}} = \max_{k,j=1,\ldots, d_k} \left| \eta_k \right|^{m-j} \left( \binom{m}{j} \right),
\] (96)

where \( J_k \) is the \( k \)th Jordan block of \( J \) with eigenvalue \( \eta_k \) and dimension \( d_k \). By proposition 16, \((\Lambda^\otimes 2)^C\) has at most one eigenvalue of modulus 1, which can be identified as the top left entry in the expression in equation (89). Consequently, all the \( \eta_k \) have modulus strictly less than 1 and so \( \left\| J^m \right\|_{\text{max}} \) and consequently \( \left\| \left( \varphi^\otimes 2 \right)^C \right\|_\infty^m \) decay exponentially to zero with \( m \).

Therefore the only term in equation (47) that does not decay exponentially to zero in \( m \) is the first term, namely
\[
\rho_0^2 E^\otimes 2 P_1 \alpha^\otimes 2 \sum_{r=1}^{m} \lambda_r^i \rho,
\] (97)

which converges exponentially in \( m \) to
\[
\rho_0^2 E^\otimes 2 P_1 \alpha^\otimes 2 \frac{1 - \lambda_1}{1 - \lambda_1},
\] (98)

provided \( |\lambda_i| < 1 \), otherwise it diverges. Since \( P_1 = uu^\dagger \) for some unit vector \( u \) and the trivial rep occurs with multiplicity 1, \( u \) is an eigenvector of \((\varphi^\otimes 2)^C\) with eigenvalue \( \lambda_1 \), which must be strictly less than 1 by proposition 16.

\[ \square \]

8. Stability under gate-dependent perturbations

In our treatment of randomized benchmarking, we have assumed that the noise is independent of the target gate (although the noise may depend on time). In a physical implementation, the noise will depend on the target. We can account for gate-dependent noise perturbatively by
writing
\[ \Lambda_{t,g} = \Lambda_t + \epsilon \Delta_{t,g}, \]  
(99)
where \( \Lambda_t = |G|^{-1} \sum_{g \in G} \Lambda_{t,g} \) is a valid quantum channel and \( \epsilon \) is scaled such that \( \| \Delta_{t,g} \|_\infty \leq 1 \) for all \( t \) and \( g \).

In [10] it was shown that the mean of the benchmarking distribution is robust under gate-dependent perturbations. We now show that the variance is also stable under gate-dependent perturbations. Let us write \( \sigma_{m,0}^2 \) for the gate-averaged variance, and \( \delta(\sigma_{m,0}^2) \) as the correction due gate-dependent perturbations. Then we have the following theorem.

**Theorem 18.** The gate-dependent correction to the variance satisfies \( \delta(\sigma_{m,0}^2) \leq \delta_0 \) whenever the gate-dependent noise in equation (99) satisfies
\[ \epsilon \leq \frac{\delta_0}{9dm}. \]  
(100)

**Proof.** The variance can be written in terms of an average over all gate sequences of length \( m \) as
\[ \sigma_{m,0}^2 = -F_m^2 + |G|^{-m} \sum_k F_{m,k}^2 \]
\[ = -F_m^2 - |G|^{-m} \left[ \sum_k F_{m,k,0}^2 + \delta\left(F_{m,k}ight) \right] \]
\[ \leq \sigma_{m,0}^2 + 2 |G|^{-m} \left| \sum_k \delta\left(F_{m,k}\right) \right| \]
\[ \leq \sigma_{m,0}^2 + 4 |G|^{-m} \left| \sum_k \delta\left(F_{m,k}\right) \right|. \]  
(101)
where \( M_0 \) and \( \delta(M) \) denote the average term and the perturbation from the average of \( M \) respectively and we have used
\[ |\delta(M^2)| = \left| \left[ M_0 + \delta(M) \right]^2 - M_0^2 \right| = 2 |M_0 + \delta(M)| |\delta(M)| \leq 2 |\delta(M)|. \]  
(102)
The second-to-last inequality follows since \( M_0 \) and \( M_0 + \delta(M) \) are in the unit interval. We have also used \( |\delta(F_{m,k})| \leq |G|^{-1} \sum_{g \in G} |F_{m,k}| \) which follows from the triangle inequality.

With noise written in the form of equation (99), the sequence of operators applied in the randomized benchmarking experiment with sequence \( S_{m,k} \) is
\[ S_{m,k} = \prod_{t=m}^0 g_t \Lambda_{t,g} = \sum_{a=0}^{m+1} \epsilon^a \sum_{b \in \mathbb{Z}_2^m} \prod_{t=m}^0 g_t M_{t,g,b_i}, \]  
(103)
where \( H(b) \) is the Hamming weight of the bit string \( b \) and
\[ M_{t,g,b_i} = \begin{cases} \Lambda_t & \text{if } b_i = 0 \\ \Delta_{t,g} & \text{if } b_i = 1. \end{cases} \]  
(104)
Substituting equation (103) into the expression for the probability \( F_{m,k} = \langle E|S_{m,k}|\rho \rangle \) and using the triangle inequality gives
\[ |\delta(F_{m,k})| \leq \sum_{a=1}^{m+1} e^a \sum_b \left| \left( E \left( \Pi_{t=m}^0 g_t M_{t,g,b} \right) \rho \right) \right| \]
\[ \leq \sum_{a=1}^{m+1} e^a \sum_b \left\| \Pi_{t=m}^0 g_t M_{t,g,b} \right\|_\infty \]
\[ \leq \sum_{a=1}^{m+1} e^a \binom{m+1}{a} d^{(a+1)/2} \]
\[ = \sqrt{d} \left[ (1 + e\sqrt{d})^{m+1} - 1 \right] \]
\[ \leq \sqrt{d} \left[ e^{\sqrt{d}(m+1)} - 1 \right]. \tag{105} \]

where, to get from the second to the third line, we note that for a fixed order \( a \) there are at most \( a + 1 \) quantum channels (i.e., products of \( \Lambda_t \)'s and the elements of \( \mathcal{G} \), which are channels) interleaved by the \( a \) different \( \Delta_{t,g} \) and we have also used the submultiplicativity of the spectral norm and proposition 5.

We can then substitute the above sequence-independent upper bound into equation (101), setting \( \delta_0 = |\delta(\sigma_m^2) - \sigma_{m,0}^2| \) and solving for \( \epsilon \) gives the sufficient condition
\[ \epsilon \leq \frac{\ln \left( 1 + \frac{\delta_0}{4\sqrt{d}} \right)}{\sqrt{d}(m+1)}. \tag{106} \]

To extract the slightly weaker but more transparent bound stated in the theorem, we use the simple bounds \( m + 1 \leq 2m \) (for \( m \geq 1 \)), \( \delta_0 \leq 1/4 \) (because the fidelity is contained in the unit interval), the inequality \( x/(1 + x) \leq \log (1 + x) \), and the loose bound \( 4\sqrt{d} + \delta_0 \leq 9/2\sqrt{d} \) for \( d \geq 2 \).

We remark that this result can surely be improved, though we have not attempted to do so. In particular, there should certainly be a factor of at least \( r \), the average infidelity, bounding the change in the variance.

9. Conclusion

We have proven that the randomized benchmarking protocol can be applied to experimental scenarios in which the noise is time dependent in an efficient and reliable manner. Moreover, the ability to estimate time-dependent average gate fidelities using randomized benchmarking provides an indicator for non-Markovianity over long timescales.

In particular, we have proven that the variance is small for short sequences and asymptotically decays exponentially to a (small) constant, that, in the case of unital noise, is zero. The fact that the variance is remarkably small (e.g., on the order of \( 4 \times 10^{-4} \) for currently achievable noise levels) enables experimental realizations of randomized benchmarking to be accurate even when using a small number of random sequences (e.g., 145 sequences compared to the \( 10^5 \) proposed in [10]).

Our results show rigorously that randomized benchmarking with arbitrary Markovian noise is generically almost as accurate as has previously been estimated in experiments [11, 12, 15] and numerics [16]. However, we find that \( K_m \) should scale with \( m \) so that the
variance is independent of $m$. We also find that if near-unitary noise (such as under- and over-rotations) are a predominant noise source, then randomized benchmarking should be conducted in the regime $mr \ll 1$, since the variance due to sampling random sequences with such noise sources will remain large as $m$ increases.

It has recently been suggested that the unexpectedly good accuracy of randomized benchmarking arises because data is simultaneously fit to $\hat{F}_m$ for all sequence lengths [16]. However, this suggestion presupposed a more fundamental fact, which we have now proven, namely, that the variance due to sampling random gate sequences is remarkably small.

Our results also apply directly to interleaved benchmarking [12] since the interleaved gate sequence can be rewritten as a standard randomized benchmarking gate sequence with the interleaved gate and its inversion incorporated into the noise, with a consequent (but small) increase in the error rate. As such, interleaved randomized benchmarking is essentially as accurate as randomized benchmarking, provided the noise is Markovian and approximately gate-independent.

Another possible application of our results is in estimating the nonunitality of a channel by estimating the constant to which the variance asymptotically converges. While it is not immediately apparent how to guarantee that the variance has (approximately) converged (given that the decay rate of the variance can be arbitrarily small), it may be possible to artificially boost the decay rate using, for example, the technique introduced in [21].

While our results prove that randomized benchmarking can reliably be performed using the number of sequences currently used in practice for qubits, the weaker bound on the variance for qudits implies that, to obtain results that are currently rigorously justified to a given confidence level, many more sequences are required when benchmarking higher-dimensional systems (or multiple qubits). Consequently, a major open problem is to improve the bound for qudits. Since the primary source of the improvement for qubits arose by considering the irrep structure of the tensor product representation, one route to obtaining an improved bound for qudits is to analyze the general representation structure of tensor product representations.

Acknowledgments

JJW acknowledges helpful discussions with Joseph Emerson. The authors were supported by the IARPA MQCO program, by the ARC via EQuS project number CE11001013, and by the US Army Research Office grant numbers W911NF-14-1-0098 and W911NF-14-1-0103. STF also acknowledges support from an ARC Future Fellowship FT130101744.

References

[1] Gottesman D 2009 An introduction to quantum error correction and fault-tolerant quantum computation, arXiv:0904.2557v1
[2] Chuang I L and Nielsen M A 1997 Prescription for experimental determination of the dynamics of a quantum black box J. Mod. Opt. 44 2455
[3] Poyatos J, Cirac J and Zoller P 1997 Complete characterization of a quantum process: the two-bit quantum gate Phys. Rev. Lett. 78 390
[4] Flammia S T, Gross D, Liu Y-K and Eisert J 2012 Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators New J. Phys. 14 095022
[5] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert Jens 2010 Quantum state tomography via compressed sensing Phys. Rev. Lett. 105 150401
[6] Merkel S T, Gambetta J M, Smolin J A, Poletto S, Córcoles A D, Johnson B R, Ryan C A and Steffen M 2013 Self-consistent quantum process tomography Phys. Rev. A 87 062119
[7] Emerson J, Alicki R and Życzkowski K 2005 Scalable noise estimation with random unitary operators J. Opt. B: Quantum Semiclass. Opt. 7 S347
[8] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R, Jost J D, Langer C, Ozeri R, Seidelin S and Wineland D J 2008 Randomized benchmarking of quantum gates Phys. Rev. A 77 012307
[9] Magesan E, Gambetta J M and Emerson J 2011 Scalable and robust randomized benchmarking of quantum processes Phys. Rev. Lett. 106 180504
[10] Magesan E, Gambetta J M and Emerson J 2012 Characterizing quantum gates via randomized benchmarking Phys. Rev. A 85 042311
[11] Gaebler J P et al 2012 Randomized benchmarking of multiqubit gates Phys. Rev. Lett. 108 260503
[12] Easwar M et al 2012 Efficient measurement of quantum gate error by interleaved randomized benchmarking Phys. Rev. Lett. 109 080505
[13] da Silva M P, Landon-Cardinal O and Poulin D 2011 Practical characterization of quantum devices without tomography Phys. Rev. Lett. 107 210404
[14] Flammia S T and Liu Y-K 2011 Direct fidelity estimation from few Pauli measurements Phys. Rev. Lett. 106 230501
[15] Brown K R, Wilson A C, Colombe Y, Ospelkaus C, Meier A M, Knill E, Leibfried D and Wineland D J 2011 Single-qubit-gate error below $10^{-4}$ in a trapped ion Phys. Rev. A 84 030303
[16] Epstein J M, Cross A W, Magesan E and Gambetta J M 2014 Investigating the limits of randomized benchmarking protocols Phys. Rev. A 89 062321 (arXiv:1308.2928v1)
[17] Granade C, Ferrie C and Cory D G 2014 Accelerated randomized benchmarking, arXiv:1404.5275
[18] Dankert C, Cleve R, Emerson J and Livine E 2009 Exact and approximate unitary two-designs and their application to fidelity estimation Phys. Rev. A 80 012304
[19] Gross D, Audenaert K and Eisert J 2007 Evenly distributed unitaries: on the structure of unitary designs J. Math. Phys. 48 052104
[20] Nielsen M A 2002 A simple formula for the average gate fidelity of a quantum dynamical operation Phys. Lett. A 303 249
[21] Kimmel S, da Silva M P, Ryan C A, Johnson B R and Ohki T A 2013 Robust extraction of tomographic information via randomized benchmarking Phys. Rev. A 87 011050 (arXiv:1306.2348)
[22] Gambetta J M et al 2012 Characterization of addressability by simultaneous randomized benchmarking Phys. Rev. Lett. 109 240504
[23] Hoeffding W 1963 Probability inequalities for sums of bounded random variables J. Am. Stat. Assoc. 58 13
[24] Ruskai M B, Szarek S and Werner E 2002 An analysis of completely-positive trace-preserving maps on Linear Algebr. Appl. 347 159
[25] Johnson B R, Ryan C A, da Silva M P, Kimmel S and Ohki T 2014 in preparation
[26] Goodman R and Wallach N R 2009 Symmetry, Representations, and Invariants (Graduate Texts in Mathematics) (Berlin: Springer)
[27] Wolf M M and Cirac J I 2008 Dividing quantum channels Commun. Math. Phys. 279 147–68
[28] Pérez-García D, Wolf M M, Petz D and Ruskai M B 2006 Contractivity of positive and trace-preserving maps under $L_p$ norms J. Math. Phys. 47 083506
[29] Evans D E and Krohn R H 1978 Probability inequalities for sums of bounded random variables J. Am. Stat. Assoc. 58 13
[30] Fuchs C A and Graaf J V D 1999 Spectral properties of positive maps on $C^*$-algebras J. London Math. Soc. 17 345–55
[31] Gilchrist A, Langford N and Nielsen M 2005 Distance measures to compare real and ideal quantum processes Phys. Rev. A 71 062310
[32] Kitaev A Y 1997 Quantum computations: algorithms and error correction *Russ. Math. Surv.* **52** 1191–249
[33] Horodecki M, Horodecki P and Horodecki R 1999 General teleportation channel, singlet fraction, and quasidistillation *Phys. Rev. A* **60** 1888
[34] Watrous J 2012 Simpler semidefinite programs for completely bounded norms, arXiv:1207.5726v2
[35] Bhatia R 1997 *Matrix Analysis* (Berlin: Springer)
[36] Blume-Kohout R, Gamble J K, Nielsen E, Mizrahi J, Sterk J D and Maunz P 2013 Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, arXiv:1310.4492v1
[37] Mirsky L 1975 A trace inequality of John von Neumann *Monatshefte Math.* **79** 303–6
[38] Sanz M, Pérez-García D, Wolf M M and Cirac J I 2010 A quantum version of Wielandt’s inequality *IEEE Trans. Inf. Theory* **56** 4668–73