Randomized Benchmarking Beyond Groups

Jianxin Chen\textsuperscript{1}, Dawei Ding\textsuperscript{2}, and Cupjin Huang\textsuperscript{1}

\textsuperscript{1}Alibaba Quantum Laboratory, Alibaba Group USA, Bellevue, Washington 98004, USA
\textsuperscript{2}Alibaba Quantum Laboratory, Alibaba Group USA, Sunnyvale, California 94085, USA

Abstract

Randomized benchmarking (RB) is the gold standard for experimentally evaluating the quality of quantum operations. The current framework for RB is centered on groups and their representations, but this can be problematic. For example, Clifford circuits need up to $O(n^2)$ gates, and thus Clifford RB cannot scale to larger devices. Attempts to remedy this include new schemes such as linear cross-entropy benchmarking (XEB), cycle benchmarking, and non-uniform RB, but they do not fall within the group-based RB framework. In this work, we formulate the universal randomized benchmarking (URB) framework which does away with the group structure and also replaces the recovery gate plus measurement component with a general “post-processing” POVM. Not only does this framework cover most of the existing benchmarking schemes, but it also gives the language for and helps inspire the formulation of new schemes. We specifically consider a class of URB schemes called twirling schemes. For twirling schemes, the post-processing POVM approximately factorizes into an intermediate channel, inverting maps, and a final measurement. This leads us to study the twirling map corresponding to the gate ensemble specified by the scheme. We prove that if this twirling map is strictly within unit distance of the Haar twirling map in induced diamond norm, the probability of measurement as a function of gate length is a single exponential decay up to small error terms. The core technical tool we use is the matrix perturbation theory of linear operators on quantum channels.

Contents

1 Introduction ................................................................. 2

2 Preliminaries ................................................................. 4

2.1 Linear Operators and Norms ......................................... 4

2.2 Matrix Perturbation Theory ......................................... 7

3 The URB Framework ...................................................... 9

3.1 URB Scheme and Experiment ....................................... 9

3.2 Examples of URB Schemes ........................................ 12

4 Main Result ................................................................. 15

4.1 Proof of Exponential Decay under Approximate Twirl ............. 15

4.2 Proving Measures are Approximate Twirls .......................... 19

4.3 Approximate Twirls in Other Norms .............................. 22
1 Introduction

With recent breakthroughs in the hardware development of quantum processors, it becomes increasingly important to be able to efficiently characterize their performance. Such a task, often called benchmarking, is essential in directing future quantum hardware development. An efficient and reliable benchmarking scheme not only allows comparison between different physical platforms and designs, but also provides useful feedback for device calibration and error diagnosis. This in turn provides useful information for future hardware designs, and, eventually, achieving fault-tolerant quantum computing.

A large number of such benchmarking schemes is collectively called randomized benchmarking (RB). Randomized benchmarking aims to extract information about a certain gate or a collection of gates through runs of random gate sequences, while isolating the effect of state preparation and measurement (SPAM) errors. Implementing the RB experiment usually produces an exponential decay curve with respect to the length of the random gate sequences. It is widely assumed that the decay rate of this curve, which can be obtained via fitting the experimental data, indicates a certain fidelity measure of the given gate set [1]. Since its first proposal in [1], randomized benchmarking has been thoroughly analyzed for its theoretical correctness and robustness [2, 3, 4, 5, 6], and it has been extensively validated by experiments [2, 7, 8, 9, 10]. Many variants of the original RB protocol have been proposed due to its great flexibility and adaptability to different experimental scenarios [3, 11, 12, 13, 14].

One central question in the study of RB is choosing the probability distributions according to which the random gate sequences are chosen. Most RB schemes require that the gate set forms a group, as the final gate in most RB experiments inverts all the previously applied gates, hence requiring the gate set to be closed under inversion of products. However, requiring that the gate set form a group is not always feasible nor necessary in an experimental setting. We give several reasons:

- Realizing an inverse gate can be experimentally challenging. In superconducting devices, arbitrary single-qubit gates can be easily implemented with high fidelities, whereas realizing a new two-qubit gate usually requires significant additional control and calibration. As a result, a connected qubit pair usually only admits a single calibrated two-qubit gate. For two-qubit gates that are locally equivalent to their inverses (or equivalently, locally equivalent to a real rotation in $SO(4)$ [15]), inverting them only requires appending the corresponding single-qubit

---

1 In fact, most RB schemes specify that all but the final gates are drawn i.i.d. uniformly or Haar-randomly from the group.
gates. However, this is not true for all two-qubit gates. Counterexamples include the fSim
gates which are indeed used in benchmarking experiments [16].

• Group structure can be very costly to implement. A typical Haar random element in $SU(2^n)$
would require a circuit of exponential length and depth to realize. A random Clifford gate on
$n$ qubits on average requires $\Omega(n^2)$ size and $\Omega(n/\log n)$ depth [17]. Neither of these groups
can be used to extract any useful information given the noise levels on current physical devices;
the signal would be long gone after implementing just a few group elements.

• Recently, linear cross-entropy benchmarking [16] was proposed as an experiment-friendly
alternative to RB on large quantum devices with tens of qubits. Linear XEB uses random
shallow circuits whose unitaries clearly do not form a group, but an exponential decay was
nevertheless observed experimentally. In fact, multiple existing benchmarking schemes [18, 19,
20] rely on random sequences of shallow circuits whose unitaries do not form groups.

In this paper, we extend the RB framework beyond groups, and propose a universal randomized
benchmarking (URB) framework which incorporates most of existing RB schemes in particular
ones that cannot be easily formulated as group-based protocols. Most importantly, the random
gates do not need to form a group but can be any set. Furthermore, instead of a single recovery
gate that is often applied at the end of the group-based RB sequence (which is often well-defined
given the group structure), we allow for a more general post-processing POVM that depends on
the random set elements chosen. For known RB schemes, intuitively the post-processing POVM
verifies that the gate sequence was perfectly applied, which is the case for group-based RB where the
post-processing POVM factors into the recovery gate and the final measurement. However, under
our framework any gate sequence dependent POVM is allowed, even ones that do not follow the
intuition of verification. The general, possibly not a group, gate set and the general post-processing
POVM define the essence of the generality provided by the URB framework. This helps open the
doors to thinking about RB from a new angle and may motivate fundamentally different schemes.

Our main technical contribution is determining conditions under which a URB scheme experiment
gives rise to a single exponential decay. The usual strategy to prove exponential decay for an RB
scheme is to first formulate the RB measurement probability as a linear functional of powers of a
linear operator. The linear operator represents the average effect of one random gate, and the power
is the sequence length. The spectral properties of the linear operator then gives the exponential decay.
For group-based RB with gate-independent noise, this operator is the twirled noise channel [1, 3, 11].
For group-based RB with gate-dependent noise it is the Fourier transformation of the implementation
map [5, 6, 21]. For our case where a group structure is not present, we assume the post-processing
POVM involves in a certain sense inverting the random gates, leading us to study linear operators
on the set of quantum channels known as twirling maps:

$$\Lambda^*_R : \mathcal{N} \mapsto \int_{g \sim \mu} dg \omega(g)^\dagger \circ \mathcal{N} \circ \omega(g),$$

where $\mu$ is some measure over a gate set and $\omega$ is a map from the gate set to $SU(d)$. Intuitively,
these are higher-order operators representing the averaged joint action of a random gate and its
inversion. They appeared in [22] to analyze a specific scheme without group structure. We can prove

\footnote{With the exception of leakage benchmarking protocols, the correct formulation of which might require quantum
channels on infinite-dimensional Hilbert spaces.}
that single exponential decays can be observed in general when the corresponding twirling map is within $\gamma$ of the Haar twirling map (set $\mu$ as the Haar measure over $SU(d)$) in induced diamond norm with $\gamma < 1$. This does not involve any underlying group structure, but instead the matrix perturbation theory of higher-order operators.

The rest of the paper is organized as follows. We first introduce basic notation, introducing operator norms and matrix perturbation theory of higher-order operators in Section 2. We then introduce the URB framework, how an experiment would run, and state sufficient conditions to obtain a single exponential decay in Section 3, together with examples of existing benchmarking schemes formulated as URB schemes. Section 4 presents the proof of the single exponential decay under the assumptions and other relevant technical details. Section 5 concludes with a discussion and open problems.

2 Preliminaries

2.1 Linear Operators and Norms

Here we give an exposition of different norms on linear operators that we will use.

**Norms on Hermitian matrices.** The spectrum of a Hermitian matrix is a vector in $\mathbb{R}^d$. We can then define norms of a Hermitian matrix via vector $\ell_p$-norms of its spectrum. For general matrices, these are known as Schatten $p$-norms. We consider three norms:

- **Trace norm**
  \[ \|\rho\|_1 := \|\text{spec}(\rho)\|_{\ell_1} = \text{tr}|\rho|. \]

- **Frobenius norm**
  \[ \|\rho\|_F := \|\text{spec}(\rho)\|_{\ell_2} = \sqrt{\text{tr}|\rho|^2}. \]

- **Spectral norm**
  \[ \|\rho\|_\infty := \|\text{spec}(\rho)\|_{\ell_\infty} = \max_i |\lambda_i(\rho)|. \]

It is easy to see that

\[ \|\rho\|_\infty \leq \|\rho\|_F \leq \|\rho\|_1 \leq \sqrt{d} \cdot \|\rho\|_F. \]  \hspace{1cm} (2.1)

Quantum states are positive semidefinite operators with unit trace, and a positive operator-valued measurement (POVM) element is a positive semidefinite operator with spectral norm upper bounded by 1.

Define the Hilbert-Schmidt inner product on Hermitian matrices as $(\cdot, \cdot)_{HS} : (\rho, \sigma) \mapsto \text{tr}[\rho\sigma]$. It is easy to see that the Frobenius norm of Hermitian matrices is induced by this inner product.

**Norms on real superoperators** Consider linear maps $\mathcal{C}$ on Hermitian matrices, which we call *real superoperators*. We can treat them as usual linear maps and define the corresponding operator norms:

- **Induced trace norm**
  \[ \|\mathcal{C}\|_{\text{tr}} := \max_{\rho \neq 0} \frac{\|\mathcal{C}(\rho)\|_1}{\|\rho\|_1}. \]
• Induced Frobenius norm

\[ \|C\|_2 := \max_{\rho \neq 0} \frac{\|C(\rho)\|_F}{\|\rho\|_F}. \]

We can also define another norm via the superoperator inner product:

\[ \langle \cdot, \cdot \rangle_{SO} : (C, D) \mapsto \sum_i \langle C(X_i), D(X_i) \rangle_{HS}, \]

where \( \{X_i\} \) is an orthonormal basis of matrices under the Hilbert-Schmidt inner product. Note that \( \{X_i\} \) can be any orthonormal basis of matrices, and we can extend the action of \( C \) to non-Hermitian matrices via linearity. We can then define

\[ \|C\|_{SO} := \sqrt{\langle C, C \rangle_{SO}}. \]

This norm is the analogue of the Frobenius norm for matrices. Another norm, called the diamond norm, is defined on composite systems:

\[ \|C\|_\diamond := \max_{\rho \neq 0} \frac{\|\rho \otimes \text{id} \|_F}{\|\rho\|_F}, \]

where \( \text{id} \) is the identity map on \( \mathbb{C}^{d \times d} \).

From the definitions and Equation (2.1) we have the following relations between the norms:

\[ \|C\|_\text{tr} \leq \|C\|_\diamond, \]

\[ \sqrt{d^{-1}}\|C\|_2 \leq \|C\|_\text{tr} \leq \sqrt{d}\|C\|_2, \]

\[ d^{-1}\|C\|_2 \leq \|C\|_\diamond \leq d\|C\|_2. \quad (2.2) \]

We can also prove the following norm inequality:

\[ \|C\|_{SO}^2 = \sum_{i,j} \|C(E(i,j))\|_F^2 \]

\[ \leq \sum_{i,j} \|C(E(i,j))\|_1^2 \]

\[ = \sum_{i,j} \|C(E(i,j)_H + E(i,j)_{AH})\|_1^2 \]

\[ \leq \sum_{i,j} \|C(E(i,j)_H)\|_1^2 + 2\|C(E(i,j)_H)\|_1 \|iC(E(i,j)_{AH}/i)\|_1 + \|iC(E(i,j)_{AH}/i)\|_1^2 \]

\[ \leq 4d^2\|C\|_{tr}^2, \]

where \( E(i,j)_{kl} := \delta_{ik}\delta_{jl} \) are the elementary matrices, which are orthonormal under the Hilbert-Schmidt inner product and have unit trace norms, and \( H \) and \( AH \) denote the Hermitian and anti-Hermitian parts, respectively. We conclude

\[ \|C\|_{SO} \leq 2d\|C\|_{tr}. \quad (2.3) \]
For the other direction, we argue
\[ \|C\|_\text{tr} \leq \sqrt{d}\|C\|_2 \leq \sqrt{d}\|C\|_{SO}, \]
where the second inequality follows since any Hermitian matrix can be extended to a basis.

A quantum channel is a completely positive, trace-preserving (CPTP) map, and consequently has unit induced trace norm and unit diamond norm. Furthermore, a quantum channel has unit induced Frobenius norm if it is a unitary, and sub-unit induced Frobenius norm if it is a mixture of unitaries. Throughout the paper we will mainly consider the Hilbert space spanned by all channels equipped with the \( \langle \cdot, \cdot \rangle_{\text{SO}} \) inner product, which we denote as \( V(d) \). There is a subspace \( V_0(d) \) with codimension 1 spanned by all differences of quantum channels. This with an arbitrary quantum channel spans the whole \( V(d) \). All norms defined on real superoperators naturally carries to \( V(d) \) and \( V_0(d) \).

It can be verified that the above norms are all bona fide matrix norms, namely
\[ \|C + D\| \leq \|C\| + \|D\|, \quad \|C \circ D\| \leq \|C\|\|D\| \]
for arbitrary \( C \) and \( D \). Furthermore, for the SO norm and the induced Frobenius norm we have
\[ \|C \circ D\|_{SO} \leq \|C\|_2\|D\|_{SO}, \quad \|C\|_{SO}\|D\|_2. \]  
(2.4)

**Norms on linear maps on real superoperators** In this work, we investigate linear operators on real superoperators, which we call *twirling maps*. Again, we can treat them like linear operators. One can define the *induced diamond norm* of a twirling map \( \Lambda \) as
\[ \|||\Lambda|||_\Diamond := \max_{C \neq 0} \frac{\|\Lambda(C)\|_\Diamond}{\|C\|_\Diamond}. \]
We can similarly define the induced SO norm \( \||| \cdot |||_2 := \| \cdot \|_{SO \rightarrow SO} \) and induced trace norm \( \||| \cdot |||_{\text{tr}} := \| \cdot \|_{\text{tr} \rightarrow \text{tr}} \), where to avoid awkwardness, we omitted the second “induced”. Again, these are all bona fide matrix norms. Note that the \( \||| \cdot |||_2 \) norm corresponds to the usual spectral norm under the superoperator inner product. Although we do not use them here, a detailed discussion involving norms and approximate twirls and unitary designs is given in \[23\]. We have the following relation similar to Equation (2.2):
\[ d^{-3/2}\|||\Lambda|||_2 \leq \|||\Lambda|||_{\text{tr}} \leq d^{3/2}\|||\Lambda|||_2. \]  
(2.5)

Moreover, twirling maps have the following property.

**Proposition 1** (Data Processing Inequality). If a twirling map \( \Lambda \) can be decomposed as
\[ \Lambda(\cdot) = \int_g dg C(g) \circ \circ D(g), \]
then
\[ \|||\Lambda||| \leq \int_g dg \|C(g)\||\|D(g)\|. \]

---

\(^3\)Without specifying otherwise, all norms are induced from real superoperators. Note that restricting to subspaces \( V(d) \) or \( V_0(d) \) does not change the inequalities and does not increase the induced norms.
where we have the appropriate correspondence between twirling map and real superoperator norms. Moreover, we have a tighter bound regarding the induced SO norm on twirling maps and induced Frobenius norm on real superoperators:

\[
|||\Lambda|||_2 \leq \int g \|C(g)\|_2 \|D(g)\|_2.
\]

**Proof.** The first inequality follows from the triangle inequality plus submultiplicativity, while the second inequality follows from the triangle inequality and Equation (2.4).

### 2.2 Matrix Perturbation Theory

We first state a few results regarding the perturbation of matrices. Let \( \mathcal{H} \) be a finite dimensional Hilbert space and \( V_1, V_2 \) be subspaces such that \( V_1 \oplus V_2 = \mathcal{H} \). Let \( A_1 \) and \( A_2 \) be operators on \( V_1 \) and \( V_2 \) respectively. Let \( \mathcal{M}(V_2, V_1) \) be the set of linear operators from \( V_2 \) to \( V_1 \), that is, linear operators \( P = X_1 P X_2 \) where \( X_1 \) and \( X_2 \) are the projectors onto \( V_1 \) and \( V_2 \) respectively. For a norm \( \| \cdot \| \) defined on \( \mathcal{H} \), we define the corresponding separation function \( \text{sep}(\cdot, \cdot) \) as

\[
\text{sep}(A_1, A_2) := \inf_{P \in \mathcal{M}(V_2, V_1)} \left\{ 0 \right\} \frac{\|A_1 P - PA_2\|}{\|P\|}.
\]

From \[24\] we have the following result.

**Theorem 2** (Stewart and Sun [24]). Let \( \mathcal{H} \) be a finite-dimensional Hilbert space with norm \( \| \cdot \| \). This norm naturally induces a norm on linear operators. Let \( V_1 \) be a linear subspace of \( \mathcal{H} \) and \( V_2 \) be its orthogonal complement. Let \( X_1 \) and \( X_2 \) be the projectors onto \( V_1 \) and \( V_2 \) respectively. Let \( A \) be a linear operator on \( \mathcal{H} \) such that

\[
X_i AX_j = A_i \delta_{ij}, \quad i, j \in \{1, 2\}.
\]

Let \( E \) be an arbitrary operator. If \( E \) satisfies

\[
\text{sep}(A_1, A_2) - \|X_1 EX_1\| - \|X_2 EX_2\| > 0,
\]

\[
\frac{\|X_1 EX_2\| \|X_2 EX_1\|}{(\text{sep}(A_1, A_2) - \|X_1 EX_1\| - \|X_2 EX_2\|)^2} < \frac{1}{4},
\]

then there exist operators \( P_1, P_2 \) such that

\[
X_2 P_1 X_1 = P_1, \quad X_1 P_2 X_2 = P_2,
\]

\[
\|P_1\| \leq \frac{2\|X_2 EX_1\|}{\text{sep}(A_1, A_2) - \|X_1 EX_1\| - \|X_2 EX_2\|},
\]

\[
\|P_2\| \leq \frac{\|X_2 EX_1\|}{\text{sep}(A_1, A_2) - \|X_1 EX_1\| - \|X_2 EX_2\| - 2\|P_1\| \|X_1 EX_2\|},
\]

such that \( A + E \) can be diagonalized as

\[
L_i^j (A + E) R_j = A'_i \delta_{ij},
\]
where
\[
L_i^1 = \begin{pmatrix} I & -P_2 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ -P_1 & I \end{pmatrix} X_i^\dagger,
\]
\[
R_i = X_i \begin{pmatrix} I & 0 \\ P_1 & I \end{pmatrix} \begin{pmatrix} I & P_2 \\ 0 & I \end{pmatrix}, i \in \{1, 2\},
\]
and
\[
A_1' = A_1 + X_1 E X_1 + X_1 E P_1,
\]
\[
A_2' = A_2 + X_2 E X_2 + P_1 E X_2.
\]

We will actually make use of a corollary that has stronger assumptions:

**Corollary 3.** Let \( \mathcal{H} \) be a finite-dimensional Hilbert space with norm \( \| \cdot \| \). This norm naturally induces a norm on linear operators. Let \( X_1, A, X_2 = I - X_1 \) be linear operators on \( \mathcal{H} \) such that
1. \( X_1 \) is an orthogonal projector,
2. \( X_i A X_j = A_i \delta_{ij}, i, j \in \{1, 2\} \),
3. \( \| A_2 \| = \gamma < 1 \),
4. \( \| E \| \leq \delta \),
5. \( A_1 = X_1 \),
6. \( \| X_1 \| \leq 1 \),
7. \( \delta \leq \frac{1 - \gamma}{3} \).

Then we obtain the conclusion of Theorem 2 with \( \| P_1 \| \leq 1, \| P_2 \| \leq 1 \). Furthermore,

- All eigenvalues of \( A_1' \) is 2\( \delta \)-close to 1,
- \( \| A_2' \| \leq (\gamma + 6\delta) \),
- \( \kappa := \| L_2 \| \| R_2 \| \leq 16 \).

**Proof.** We first prove \( \text{sep}(A_1, A_2) \geq 1 - \gamma \). This is because \( A_1 P = X_1 P = P \) and thus
\[
\text{sep}(A_1, A_2) := \inf_{P \in M(V_2, V_1), \| P \| = 1} 1 - \| PA_2 \| \geq 1 - \gamma.
\]
Furthermore, \( \| X_2 \| = \| I - X_1 \| \leq 2 \). Therefore \( \| X_1 E X_1 \| \leq \delta, \| X_2 E X_1 \| \leq 2\delta, \| X_1 E X_2 \| \leq 2\delta \) and \( \| X_2 E X_2 \| \leq 4\delta \). It is easy to verify that both assumptions in Theorem 2 hold, and
\[
\| P_1 \| \leq \frac{4\delta}{11\delta - 5\delta} \leq 1,
\]
\[
\| P_2 \| \leq \frac{2\delta}{11\delta - 5\delta - 4\delta} \leq 1.
\]
We have
\[
\| A_1' - A_1 \| \leq \| X_1 \| \| E \| (\| X_1 \| + \| P_1 \|) \leq 2\delta,
\]
\[
\| A_2' - A_2 \| \leq \| X_2 \| \| E \| (\| X_2 \| + \| P_1 \|) \leq 6\delta \Rightarrow \| A_2' \| \leq \| A_2 \| + 6\delta \leq \gamma + 6\delta.
\]
From Theorem 2 we have \( A_1' = X_1 A_1' X_1 \). Therefore for any eigenvector \( v \) of \( A_1' \) with eigenvalue \( \Lambda \), we have \( A_1 v = v \). Therefore
\[
\| (A_1 - A_1') v \| = |1 - \lambda| \| v \| \leq \| A_1 - A_1' \| \| v \| \Rightarrow |1 - \lambda| \leq \| A_1 - A_1' \| \leq 2\delta.
\]
Finally,
\[
\| L_2 \| = \| - P_1 X_1 + X_2 \| \leq 3,
\]
\[
\| R_2 \| = \| X_1 P_2 + X_2 P_1 P_2 + X_2 \| \leq 5,
\]
proving that \( \kappa \leq 16 \).

Note that assumption (1) of Corollary 3 implies the setting of Theorem 2.
3 The URB Framework

3.1 URB Scheme and Experiment

Let $\mathcal{C}(d)$ be the set of qudit channels and $H(d)$ be the set of $d \times d$ Hermitian matrices.

**Definition 4 (URB scheme).** A universal randomized benchmarking (URB) scheme $R$ on a $d$-dimensional quantum system can be expressed as a tuple $(S, \mu, \phi, M, \rho_0)$ consisting of:

- A gate set $S$, encoding the gates to be applied in the URB scheme. For sake of generality, we do not restrict $S$ to be a subset of the unitary group $SU(d)$; instead $S$ can encode any description that leads to an implementation of the gate independent of other gates in a circuit.
- A probability distribution $\mu$ over the gate set $S$.
- An implementation map $\phi: S \to \mathcal{C}(d)$, assuming a gate-dependent yet Markovian noise model.
- A post-processing POVM $M: S^* \to H(d)$ taking a finite string of elements from the set to a Hermitian operator on $\mathbb{C}^d$.
- An initial state $\rho_0 \in H(d)$.

**Definition 5 (URB experiment).** A URB scheme $R = (S, \mu, \phi, M, \rho_0)$ gives rise to the following experiment protocol:

1. Given sequence length $m$, choose $m$ random elements $g_1, \cdots, g_m \in S$ i.i.d. according to the probability distribution $\mu$.
2. Apply $\phi(g_1), \cdots, \phi(g_m)$ sequentially on a prepared initial state $\rho_0$.
3. Perform a measurement $M(g_1, \cdots, g_m)$ on the final state and get a binary result.
4. Repeat steps 1 to 3 to get an estimation $\hat{p}(m)$ of the success probability.
5. Repeat steps 1 to 4 over appropriately chosen length parameters $m_1, \cdots, m_k$ to get estimations $\hat{p}(m_1), \cdots, \hat{p}(m_k)$. Return the estimations.

It is easy to see that each $\hat{p}(m)$ is an unbiased estimator of the quantity

$$p_R(m) := \mathbb{E}_{g_1,\cdots,g_m \sim \mu} \tr[M(g_1,\cdots,g_m) \cdot (\phi(g_m) \circ \cdots \circ \phi(g_1))\rho_0].$$

The output estimations $\hat{p}(m_1), \cdots, \hat{p}(m_k)$ are in practice assumed to be sufficiently close to $p_R(m_1), \cdots, p_R(m_k)$ and are fit to a single exponential decay curve $f(m) = A + B \cdot u^m$ to extract the decay rate $u$. The decay rate $u$ is commonly believed to indicate an “average fidelity” of the gate ensemble (see Section [4.4](#) for the interpretation of the RB value). However, the above URB framework itself does not guarantee a single exponential decay. To guarantee such a decay, we need the following.

**Definition 6.** A URB scheme is called an $(\epsilon, \delta, \gamma)$ twirling scheme if the following hold:
Approximate factoring of the post-processing POVM into a triple \((M_0, \phi^*, \mathcal{I})\): a final measurement \(M_0 \in H(d)\), an inverting map \(\phi^* : S \rightarrow \mathcal{C}(d)\) and an intermediate channel \(\mathcal{I} \in \mathcal{C}(d)\) such that the post-processing POVM can be approximately factored into three parts:

\[
\sup_{m \in \mathbb{N}} \mathbb{E}_{g_1, \ldots, g_m \sim \mu} \left| \left| M(g_1, \cdots, g_m) - M_0 \cdot \phi^*(g_1) \circ \cdots \circ \phi^*(g_m) \circ \mathcal{I} \right| \right|_\infty \leq \epsilon,
\]

where we use the notation for a Hermitian matrix \(\rho\) and real superoperator \(\mathcal{C}\),

\[
\rho \cdot \mathcal{C} := \mathcal{C}^\dagger (\rho),
\]

\(\mathcal{C}^\dagger\) being the adjoint superoperator with respect to the Hilbert-Schmidt norm.\(^4\)

**Near-ideal implementation** There exists an ideal map \(\omega : S \rightarrow \mathcal{C}(d)\) into unitary channels, such that the implementation map is close to the ideal map and the inverting map is close to its adjoint:

\[
\mathbb{E}_{g \sim \mu} \left[ \left| \left| \phi(g) - \omega(g) \right| \right|_\diamond + \left| \left| \phi^*(g) - \omega(g)^\dagger \right| \right|_\diamond \right] \leq \delta.
\]

**Approximate twirling** The twirling map

\[
\Lambda^*_R : \mathcal{N} \mapsto \int_{g \sim \mu} dg \omega(g)^\dagger \circ \mathcal{N} \circ \omega(g)
\]

is a \(\gamma\)-approximate twirl: that

\[
\left| \left| \Lambda^*_R - \Lambda^* \right| \right| \leq \gamma
\]

under a certain norm \(\left| \left| \cdot \right| \right|\) (usually taken as the diamond norm) where \(\Lambda^* : \mathcal{N} \mapsto \int_{g \sim \eta} dg \tilde{g}^\dagger \circ \mathcal{N} \circ \tilde{g}\) is the twirling map where \(\eta\) is the Haar random distribution on \(SU(d)\), and \(\tilde{g} : \rho \mapsto g\rho g^{-1}\) is the unitary channel corresponding to the fundamental representation of \(SU(d)\) for \(g\).

We will also say that a URB scheme \(R\) is an \((\epsilon, \delta, \gamma)\) twirling scheme with respect to the tuple \((M_0, \phi^*, \mathcal{I}, \omega)\) in cases where the components are to be specified. The approximately factorizable post-processing POVM condition expresses that a twirling scheme is an RB-like scheme in that after the random gates, a series of inverting gates are applied followed by a measurement. However, this mathematical definition can be more inclusive than it may seem, as we will see that linear XEB also effectively uses a factorizable post-processing POVM. The near-ideal implementation condition simply expresses that the noise levels are bounded. The most interesting condition is the \(\gamma\)-approximate twirl, which expresses that our random gate distribution is constant distance from a unitary 2-design. Figure\(^5\) visually illustrates the URB framework and twirling schemes.

The main technical contribution of our work is a characterization of the exponential decay behavior provided that the URB scheme parameters satisfy certain constraints. More specifically we have the following main result.

**Theorem 7** (Theorem\(^8\) informal). Let \(R\) be an \((\epsilon, \delta, \gamma)\) twirling scheme with respect to the diamond norm, and assume that \(\gamma \leq 1 - 11\delta\). Then there exists \(A, B, p \in \mathbb{R}, p \in [1 - 2\delta, 1]\) such that

\[
|p_R(m) - (A + B \cdot p^m)| \leq \epsilon + 16(\gamma + 6\delta)^m.
\]

\(^4\)We will use \(\cdot\) to denote scalar multiplication, matrix multiplication, or the above shorthand. These can be differentiated by context.
Figure 1: A visual illustration of the URB framework and twirling schemes. (a) A generic URB scheme. Given a sequence length $m$, we randomly select $m$ i.i.d. samples from a distribution $\mu$. The experiment then applies $g_1, \cdots, g_m$ sequentially via an implementation map $\phi$ on a fixed initial state $\rho_0$, followed by a post-processing POVM $M$ that depends on the random elements $g_1, \cdots, g_m$. (b) Approximate factoring. When the post-processing POVM admits an approximate factoring, it can be approximately decomposed into an intermediate channel $\mathcal{I}$, a sequence of inverting operations $\phi^*(g_m), \cdots, \phi^*(g_1)$, and a fixed measurement operator $M_0$. (c) A twirling map. The average joint operation of $\phi(g)$ and $\phi^*(g)$ sandwiching a channel can be formulated as a linear operator $\Lambda_R$ on quantum channels. (d) A URB scheme in terms of twirling maps. Expressing the probability of measurement in terms of the twirling map $\Lambda_R$ reduces the proof of the exponential decay to the study of its spectral properties.
We see that the crucial property we need to establish to apply Theorem 7 is $\gamma < 1$, so that for sufficiently low experimental error $\delta$, $\gamma \leq 1 - 11\delta$. One may mistakenly think that our result is essentially saying there is a single exponential decay if our twirling map is close to the Haar twirl $\Lambda^*$, which intuitively means the unitary ensemble defined by $\mu, \omega$ is close to a unitary 2-design. This is of course unsurprising. However, we stress that we do not require $\gamma$ to be small, but just less than 1. This is not as strong of a requirement as being close to a unitary 2-design.

Note also that Theorem 7 by itself is not sufficient to imply we can extract a single exponential decay from measured data. In general, we require the magnitude of $B$ to be significantly larger than 0, $p$ to be significantly larger than $\gamma + 6\delta$, and $\epsilon$ is sufficiently small for the URB experiment to be able to extract a single exponential decay with decay rate close to $p$ given sufficiently many repeated experiments. For the effect of $\epsilon$, see Section 4.5 for an analysis of the robustness of fitting to a perturbed exponential decay.

3.2 Examples of URB Schemes

In the URB framework, the post-processing POVM $M$ is defined abstractly for generality. To our knowledge, all known URB schemes can be approximately factorized, but we leave open other possibilities. We here give examples of schemes that fall into our framework. Note that we make the distinction between a scheme falling into our framework and it being guaranteed by our theorem to have a single exponential decay, which requires additional assumptions. Further assumptions are required for this exponential decay to be extractable. We summarize the relationship between the URB framework, our class of twirling schemes, and the existing group-based framework as well as other classes of RB schemes in Figure 2.

Group-based Randomized Benchmarking  Standard group-based RB can be readily formulated as a URB scheme, where the gate set is taken as a group $G$, with the distribution being the uniform distribution over the group. The post-processing POVM is then $M(g_1, \cdots, g_m) = M_0 \cdot \phi(g_1^{-1} \cdots g_m^{-1})$, i.e. applying a fixed measurement after physically applying the gate corresponding to the inverse of the product of the previous elements. This post-processing POVM admits a $\delta$-approximate factoring into the triple $(M_0, \tilde{\omega}^\dagger, \text{id})$ the when the implementation map $\phi$ is $\delta$-close to a representation $\tilde{\omega}$, that is,

$$\mathbb{E}_{g \sim \mu} \| \phi(g) - \tilde{\omega}(g) \|_\diamond \leq \delta.$$  

Moreover, in the case that the uniform distribution over the image of the ideal map forms a unitary 2-design, the twirling map is an exact twirl, and a single exponential decay occurs whenever $\delta < \frac{1}{11}$. More general cases where there are different representations or multiplicities of representations still lie in the URB framework, but we do not consider them below.

Note that unlike the results for example in [6], we have an additional constant error term $\epsilon$. This is because we want to reduce our analysis to twirling maps, and to do this we need to consider $\phi^*(g_1) \circ \cdots \circ \phi^*(g_m)$ instead of $\phi((g_m \cdots g_1)^{-1})$. We also cannot consider the latter option because we lack group structure.

Non-uniform RB  There have been several extensions to standard RB [2, 26, 19] where the random gates are drawn from non-uniform distributions over a group, since a typical Haar random element is too costly. Such RB variants readily fit into our URB framework as it makes no assumption
on the distribution. In fact, URB schemes with an ideal reference map $\omega$ naturally gives rise to a (not necessarily uniform) distribution on the special unitary group of corresponding dimension.

Our work extends [26] in several ways, but most importantly we do not require the distribution to be approximately uniform, nor inverse-symmetric, nor that its support contains the generators of a group. To be a twirling scheme, we only require that the twirling map is close to Haar under a certain norm. As a result, our result provides tighter bounds with less assumptions on the distribution, and also applies to infinite groups. We leave it to future work to study a twirling scheme where there are multiple irreducible representations or irreducible representations with multiplicities, which would result in a matrix exponential decay.

**Linear Cross-entropy Benchmarking (Linear XEB)** Linear cross-entropy benchmarking (Linear XEB) is a benchmarking scheme first introduced in [16]. In this framework, a certain number of layers, typically shallow, of random circuits $C_1, \cdots, C_m$ are applied to an initial state. A subsequent measurement then returns a bitstring $x$, whose ideal probability $q(x)$ is numerically simulated on a classical computer. The protocol returns the value $2^n q(x) - 1$, where $n$ is the number of qubits in the system.

We claim that linear XEB falls into the URB framework, provided that the random shallow circuits are chosen i.i.d. from a distribution $\mu$ over a set $S$. To see this, we express the expected outcome

$$F(m) = \mathbb{E}_{C_1, \cdots, C_m \sim \mu} \mathbb{E}_{x \sim \tilde{q}} [2^n q(x) - 1],$$

(3.1)
where the distributions \( q, \tilde{q} \) are defined as

\[
q(x) = |\langle x|C_m \cdots C_1|0\rangle|^2,
\]

\[
\tilde{q}(x) = \text{tr} \left[ M_x \cdot \phi(C_m) \circ \cdots \circ \phi(C_1)(\rho_0) \right]
\]

for some POVM \( \{M_x\}_{x \in \{0,1\}^n} \) and initial state \( \rho_0 \). By lifting the distributions as diagonal operators, we have

\[
F(m) = 2^n \cdot E_{C_1,\ldots,C_m} \mu \text{tr}[Q \tilde{Q}] - 1,
\]

where

\[
Q = \sum_x q(x)|x\rangle\langle x| = D \circ C_m \circ \cdots \circ C_1(|0\rangle\langle 0|),
\]

\[
\tilde{Q} = \sum_x \tilde{q}(x)|x\rangle\langle x| = \tilde{D} \circ \phi(C_m) \circ \cdots \circ \phi(C_1)(\rho_0),
\]

with \( D : \rho \mapsto \sum_x |x\rangle\langle x| \cdot |x\rangle\langle x| \) and \( \tilde{D} : \rho \mapsto \sum_x \text{tr}[M_x \rho] \cdot |x\rangle\langle x| \) representing the ideal and physical measurement under the computational basis, and \( C_i \) the ideal implementation of \( C_i \) as a unitary channel. Using the fact that \( Q = Q^\dagger \), we have

\[
F(m) = 2^n \cdot E_{C_1,\ldots,C_m} \mu \text{tr}[|0\rangle\langle 0| \cdot C_1^\dagger \circ C_m^\dagger \circ \tilde{D} \circ \phi(C_m) \circ \cdots \circ \phi(C_1)(\rho_0)] - 1,
\]

that is, the post-processing POVM, being the combination of the quantum measurement followed by classical simulation, can be exactly factored in terms of a tuple \( (|0\rangle\langle 0|, \tilde{\omega}^\dagger, \tilde{D}) \), with a prefactor of \( 2^n \).

Linear XEB experiments suggest that \( F(m) \) can be fit into a single exponential decay \( A' + B' \cdot p^n \) with \( A' \approx 0 \) and \( B' = O(1) \). Thus, \( B = 2^{-n}B' \ll 1 \). This implies that the single exponential decay may not be observable unless the circuit depth approaches \( \omega(n \log(1 - \gamma)) \) so that the error term is negligible compared to the actual single exponential decay. We leave a more detailed analysis of linear XEB experiments for future work.

**Cycle Benchmarking** Cycle benchmarking [20] is a specialized protocol to measure the Pauli errors of Clifford gadgets in a large-scale quantum processor. Given an \( n \)-qubit Clifford gate \( g \in C_n \), let \( k = \text{ord}(g) := \min \{ k \in \mathbb{Z}_+ | g^k = I \} \). Each random element is chosen uniformly from the set of \( n \)-qubit Pauli operators \( \mathcal{P}_{m} \), and is implemented by

\[
\phi((P_1, \ldots, P_k)) := \varphi(P_1) \circ \varphi(g) \circ \cdots \circ \varphi(P_k) \circ \varphi(g),
\]

for some physical implementation \( \varphi \) defined on all Pauli gates and \( g \). It can be proven that in the noiseless case such implementations result in a Pauli operator, and it is therefore sufficient to calculate and implement the final Pauli gate as the recovery gate. The post-processing POVM can be approximately factored given \( \varphi \) on the set of Pauli operators is close to ideal.

One distinct feature of the cycle benchmarking is that there is almost never a single exponential decay: the twirling map would be far from an approximate twirl since the distribution effectively defined on the set of Pauli channels. Instead, there is typically an exponential number of exponential decay components from a single experiment. Multiple experiments, typically with different measurements, are required in order to isolate the exponential decay components to give useful information about the Pauli error channels.
4 Main Result

4.1 Proof of Exponential Decay under Approximate Twirl

Theorem 8. Let \( R = (S, \mu, \phi, M, \rho_0) \) be an \((\epsilon, \delta, \gamma)\) twirling scheme with respect to the tuple \((M_0, \phi^*, I_\omega)\). Given that \( \delta \leq \frac{1}{16} \), there exists \( A, B \in \mathbb{R} \) and \( p \in [1 - 2\delta, 1] \) such that

\[
|p_R(m) - (A + Bp^m)| \leq \epsilon + 16(\gamma + 6\delta)^m.
\]

Our proof centers around twirling maps that map channels to channels; for a URB scheme \( R = (S, \mu, \phi, M, \rho_0) \) that is a twirling scheme with parameters \((M_0, \phi^*, I_\omega)\), define the ideal twirling map

\[
\Lambda^*_R : \mathcal{N} \mapsto \int_{g \sim \mu} d\omega(g)^\dagger \circ \mathcal{N} \circ \omega(g)
\]

and the physical twirling map

\[
\Lambda_R : \mathcal{N} \mapsto \int_{g \sim \mu} d\phi^*(g) \circ \mathcal{N} \circ \phi(g).
\]

The following is a summary of the proof: We first approximate \( p_R(m) \) as a linear function of the \( m \)-th power of the physical twirling map \( \Lambda_R \), transforming the problem into the study of the major spectral components of \( \Lambda_R \). This is then analyzed by regarding \( \Lambda_R \) as a perturbed version of the ideal twirling map \( \Lambda^*_R \), whose spectral properties are given by the \( \gamma \)-approximation with respect to the Haar twirl \( \Lambda^* \). For sake of simplicity we denote the Haar random distribution \( \eta \) without explicitly specifying the dependence on the dimension \( d \).

Proof. We first compute \( p_R(m) \):

\[
p_R(m) = \mathbb{E}_{g_1, \ldots, g_m \sim \mu} \text{tr}[M(g_1, \ldots, g_m) \cdot \phi(g_m) \circ \cdots \circ \phi(g_1)(\rho_0)],
\]

Then,

\[
\begin{align*}
|p_R(m) - \text{tr}[M_0 \cdot \Lambda^*_R(I)(\rho_0)]| &= \left| \mathbb{E}_{g_1, \ldots, g_m \sim \mu} \text{tr}[M(g_1, \ldots, g_m) \cdot \phi(g_m) \circ \cdots \circ \phi(g_1)(\rho_0)] ight| \\
&\quad - \mathbb{E}_{g_1, \ldots, g_m \sim \mu} \text{tr}[M_0 \cdot \phi^*(g_1) \circ \cdots \circ \phi^*(g_m) \circ I \circ \phi(g_m) \circ \cdots \circ \phi(g_1)(\rho_0)] \\
&\leq \mathbb{E}_{g_1, \ldots, g_m \sim \mu} \left| \text{tr}[(M(g_1, \ldots, g_m) - M_0 \cdot \phi^*(g_1) \circ \cdots \circ \phi^*(g_m) \circ I)(\phi(g_m) \circ \cdots \circ \phi(g_1)(\rho_0))] \right| \\
&\leq \mathbb{E}_{g_1, \ldots, g_m \sim \mu} \|M(g_1, \ldots, g_m) - M_0 \cdot \phi^*(g_1) \circ \cdots \circ \phi^*(g_m) \circ I\|_\infty \|\phi(g_m) \circ \cdots \circ \phi(g_1)(\rho_0)\|_\text{tr} \\
&\leq \epsilon, \\
& \quad \quad (4.1)
\end{align*}
\]

\footnote{This bound can be further tightened to \( p \in [1 - \delta, 1] \) using the Bauer-Fike theorem, but we leave optimization of constants in our results for sake of conciseness.}
where the first inequality uses the triangle inequality, the second inequality uses the Hölder inequality, the third equality uses the fact that all $\phi(g_i)$’s are channel and thus $\phi(g_m) \circ \cdots \circ \phi(g_1)(\rho_0)$ is a state with unit trace, and the final inequality follows by the factorization assumption.

Now we turn to study the quantity $\text{tr}[M_0 \cdot \Lambda_R^m(I)(\rho_0)]$. We do this by studying the spectral properties of $\Lambda_R$. Specifically, we would like to invoke Corollary 3 with the following setup:

$$\mathcal{H} = V(d), A = \Lambda_R^m, X_1 = \Lambda^*, X_2 = I - \Lambda^*, E = \Lambda_R - \Lambda_R^*, \|\cdot\| = |||\cdot|||_o.$$  

We verify that all the assumptions in Corollary 3 hold.

1. $X_1^\dagger = X_1, X_2^\dagger = X_1$. The first is because for any $\mathcal{C}, \mathcal{D} \in V(d)$,

$$\langle \mathcal{C}, \Lambda^* (\mathcal{D}) \rangle_{SO} = \langle \mathcal{C}, \int_{g \sim \eta} dg \tilde{g}^\dagger \circ \mathcal{D} \circ \tilde{g} \rangle_{SO}$$

$$= \int_{g \sim \eta} dg \langle \mathcal{C}, \tilde{g}^\dagger \circ \mathcal{D} \circ \tilde{g} \rangle_{SO}$$

$$= \int_{g \sim \eta} dg \sum_i \langle \mathcal{C}(Y_i), \tilde{g}^\dagger \circ \mathcal{D} \circ \tilde{g}(Y_i) \rangle_{HS}$$

$$= \int_{g \sim \eta} dg \sum_i \langle \tilde{g} \circ \mathcal{C}(Y_i), \mathcal{D} \circ \tilde{g}(Y_i) \rangle_{HS}$$

$$= \int_{g \sim \eta} dg \sum_i \langle \tilde{g} \circ \mathcal{C} \circ \tilde{g}^\dagger(Y_i^g), \mathcal{D}(Y_i^g) \rangle_{HS}$$

$$= \int_{g \sim \eta} dg \langle \tilde{g} \circ \mathcal{C} \circ \tilde{g}^\dagger, \mathcal{D} \rangle_{SO}$$

$$= \langle \mathcal{C}, \mathcal{D} \rangle_{SO} = \langle \Lambda^*(\mathcal{C}), \mathcal{D} \rangle_{SO}.$$

Here we used the facts that $\{Y_i^g\}_i := \{\tilde{g}(Y_i)\}_i$ is an orthonormal basis under the Hilbert-Schmidt inner product if $\{Y_i\}_i$ is, that $(\tilde{g}^\dagger)^\dagger = \tilde{g}$, and that the Haar measure is inverse symmetric.

To prove $X_1^2 = X_1$, define $\Lambda(\nu) := \int_{g \sim \nu} dg \tilde{g}^\dagger \circ \tilde{g}$ for arbitrary distribution $\nu$ defined on $SU(d)$. We then have

$$\Lambda(\nu) \Lambda^* = \int_{g \sim \nu} dg \int_{h \sim \eta} dh \tilde{h}^\dagger \circ \tilde{h} \circ \tilde{g}$$

$$= \int_{g \sim \nu} dg \int_{h \sim \eta} dh (\tilde{h}g)^\dagger \cdot (\tilde{h}g)$$

$$= \int_{g \sim \nu} dg \int_{h \sim \eta} dh (g\tilde{h})^\dagger \cdot (\tilde{h}g)$$

$$= \int_{g \sim \nu} dg \int_{h' \sim \eta} dh' \tilde{h}^\dagger \cdot \tilde{h}'$$

$$= \int_{g \sim \nu} dg \Lambda^* = \Lambda^*,$$
where we have used the fact that the Haar measure is right invariant. Taking $\nu = \eta$ proves $X_1^2 = X_1$.

2. $X_i A X_j = A_i \delta_{ij}$. Since $\Lambda^*_R$ is uniquely determined by the distribution $\mu_\omega$ on $SU(d)$ which is defined by the distribution $\mu$ on $S$ and the ideal map $\omega$, we have $\Lambda^* \Lambda^*_R = \Lambda^*$ and similarly $\Lambda^*_R \Lambda^* = \Lambda^*$ from the above. This directly leads to

$$X_1 A X_1 = \Lambda^* = A_1, X_1 A X_2 = \Lambda^* \Lambda^*_R (I - \Lambda^*) = \Lambda^*(I - \Lambda^*) = \Lambda^* - \Lambda^* = 0,$$

$$X_2 A X_1 = 0, X_2 A X_2 = (I - \Lambda^*) \Lambda^*_R (I - \Lambda^*) = \Lambda^*_R - \Lambda^* = A_2.$$

3. $\|A_2\| = |||\Lambda^*_R - \Lambda^*|||_\diamond \leq \gamma$ by $\Lambda^*_R$ being a $\gamma$-approximated twirl.

4. $\|E\| \leq \delta$. This is because

$$\|E\| = |||\Lambda_R - \Lambda_R^*|||_\diamond$$

$$= |||\int_{g \sim \mu} dg (\omega(g)^{\dagger} \circ \circ \omega(g) - \phi^*(g) \circ \circ \phi(g))|||_\diamond$$

$$= |||\int_{g \sim \mu} dg ((\omega(g)^{\dagger} - \phi^*(g)) \circ \circ \omega(g) + \phi^*(g) \circ \circ (\omega(g) - \phi(g)))|||_\diamond$$

$$\leq \int_{g \sim \mu} dg |||\omega(g)^{\dagger} - \phi^*(g)|||_\diamond |||\omega(g)|||_\diamond + \int_{g \sim \mu} dg |||\phi^*(g)|||_\diamond |||\omega(g) - \phi(g)|||_\diamond$$

$$\leq \int_{g \sim \mu} dg |||\omega(g)^{\dagger} - \phi^*(g)|||_\diamond + \int_{g \sim \mu} dg |||\omega(g) - \phi(g)|||_\diamond$$

$$\leq \delta,$$

where the first inequality uses the triangle inequality and Proposition 1, and the second inequality uses that $|||\omega(g)|||_\diamond, |||\phi^*(g)|||_\diamond = 1$ since they are channels.

5. $A_1 = X_1$ is already proven above.

6. $\|X_1\|_\diamond \leq 1$ follows directly from Proposition 1.

7. This is assumed.

We can now invoke Corollary 3 to obtain $L_1, L_2, R_1, R_2, A'_1, A'_2$ such that $\Lambda_R$ is diagonalized:

$$\Lambda_R = L_1 A'_1 R_1 + L_2 A'_2 R_2,$$

and $\|A'_1 - X_1\| \leq 2\delta, \|A'_2\| \leq \gamma + 6\delta$, and $\|L_2\| \|R_2\| \leq 16$.

The diagonalization ensures that $\Lambda_R^n = L_1 (A'_1)^n R_1 + L_2 (A'_2)^n R_2$ and therefore

$$\text{tr}[M_0 \cdot \Lambda_R^n (I)(\rho_0)] = \text{tr}[M_0 \cdot (L_1 (A'_1)^m R_1 + L_2 (A'_2)^m R_2) (I)(\rho_0)]$$

$$= \text{tr}[M_0 \cdot L_1 (A'_1)^m R_1 (I)(\rho_0)] + \text{tr}[M_0 \cdot L_2 (A'_2)^m R_2 (I)(\rho_0)].$$

The second term is bounded by an exponential decay:

$$|\text{tr}[M_0 \cdot L_2 (A'_2)^m R_2 (I)(\rho_0)]| \leq \|L_2 (A'_2)^m R_2 (I)\|_\diamond$$

$$\leq \|L_2\| \cdot \|(A'_2)^m\| \cdot \|R_2\|$$

$$\leq 16 \cdot \|A'_2\|^m$$

$$\leq 16 \cdot (\gamma + 6\delta)^m.$$  (4.3)
We finally show that the term \( \text{tr}[M_0 \cdot L_1(A_1')^m R_1(I)(\rho_0)] \) is exactly a single exponential decay. Specifically, we show that the rank of \( A_1' \) is 2, and it can be diagonalized with one eigenvalue 1 and the other \( p \in [1 - 2\delta, 1] \).

1. \( A_1' \) is of rank 2. First, by construction the rank is at most 2. Also, for any quantum channel \( \mathcal{N} \),
   \[
   \Lambda^*(\mathcal{N}) = \alpha_\mathcal{N} \cdot \text{id} + (1 - \alpha_\mathcal{N}) \cdot \text{dep}
   \]
   for some \( \alpha_\mathcal{N} \). Therefore \( \text{Im}(\Lambda^*) = \text{span}\{\{\text{dep}, \text{id}\}\} \) and \( \text{rank}(\Lambda^*) = 2 \). Moreover, since \( \|A_1' - A_1\| \leq 2\delta \) and \( A_1 = \Lambda^* \) is a projector, the two eigenvalues \( \lambda_1, \lambda_2 \) of \( A_1' \) must be \( 2\delta \)-close to 1 and \( A_1' \) is therefore of rank 2.

2. **Eigenvalues of \( A_1' \) are bounded in magnitude by 1.** For any non-zero eigen-channel \( \mathcal{N} \) of \( \Lambda_R \) with eigenvalue \( \lambda \), we have
   \[
   |\lambda| \|\mathcal{N}\|_\circ = \|\Lambda_R(\mathcal{N})\|_\circ \leq \||\Lambda_R\|\|\mathcal{N}\|_\circ = |\lambda| \leq \|\Lambda_R\|_\circ \leq 1.
   \]
   Therefore all eigenvalues of \( \Lambda_R \) are bounded in magnitude by 1, and so are \( A_1', A_2' \) since their two spectra constitute a bipartition of the spectrum of \( \Lambda_R \).

3. **All \( \Lambda_R, \Lambda_R^* \) and \( \Lambda^* \) are real-valued under a properly chosen basis.** Consider an orthogonal basis \( B \) (under the Hilbert-Schmidt inner product) of the space of Hermitian matrices. Then, any real superoperator is defined by its action on elements of \( B \), giving rise to a matrix representation. In particular, real superoperators will be real-valued matrices since they preserve hermiticity. Now, we can consider a basis \( \mathcal{B} := \{\mathcal{N}_i\}_1 \) of real superoperators with a matrix representation with respect to \( B \) being an elementary matrix: a single 1 and 0's everywhere else. Since \( \omega(g), \phi(g) \) and \( \phi^*(g) \) are quantum channels for all \( g, \Lambda_R, \Lambda_R^*, \Lambda^* \) maps \( \mathcal{N}_i \) to real matrices with respect to \( B \) and therefore are themselves real matrices with respect to \( B \). This also indicates that \( L_1, L_2, R_1, R_2, A_1', A_2' \) are real with respect to \( B \). Additionally, the intermediate channel \( I \), the initial state \( \rho_0 \), and the final measurement \( M_0 \) are all real with respect to the corresponding bases.

4. **\( A_1' \) has one eigenvalue 1.** Consider the channel
   \[
   \mathcal{C}_{\phi^*} := \int_{g \sim \mu} dg \phi^*(g).
   \]
   By Brouwer’s fixed-point theorem, there exists a quantum state \( \rho^* \) such that \( \mathcal{C}_{\phi^*}(\rho^*) = \rho^* \) is an eigenstate of \( \mathcal{C}_{\phi^*} \) with eigenvalue 1. \( \rho^* \) can then be lifted to an eigen-channel \( \mathcal{E}_{\rho^*} : \rho \mapsto \rho^* \cdot \text{tr}[\rho] \) of \( \Lambda_R \) with eigenvalue 1. This eigenvalue must lie in \( A_1' \) as \( \|A_2'\| \leq \gamma + 6\delta < 1 \). Without loss of generality denote \( \lambda_1 = 1 \). Then \( \lambda_2 \) must be real as \( A_1' \) is real and so eigenvalues come in complex conjugate pairs, and \( \lambda_2 \in [1 - 2\delta, 1] \).

5. **\( A_1' \) is diagonalizable even if \( \lambda_2 = 1 \).** Suppose otherwise and \( A_1' = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \) under properly chosen \( L_1, R_1 \) with \( a \neq 0 \). Then \( \text{tr}[M_0 \cdot L_1(A_1')^m R_1(I)(\rho_0)] \) becomes a linear function with respect to \( m \) with non-zero coefficients for some \( M_0, I, \rho_0 \), and thus becomes unbounded. This contradicts the fact that it is the difference of two bounded functions:
   \[
   \text{tr}[M_0 \cdot L_1(A_1')^m R_1(I)(\rho_0)] = \text{tr}[M_0 \cdot \Lambda_R^m(I)(\rho_0)] - \text{tr}[M_0 \cdot L_2(A_2')^m R_2(I)(\rho_0)],
   \]
   where
where the magnitude of the first term is bounded by $p_R(m) + \epsilon \leq 1 + \epsilon$ and that of the second term by $16(\gamma + 6\delta)^m \leq 16$.

\[ \text{Remark 9.} \] In the proof we chose the Hilbert space $\mathcal{H}$ to be the channel space $V(d)$ as it is an invariant space under all twirling maps $\Lambda^*, \Lambda^*_R$ and $\Lambda_R$. A similar proof considers the channel difference space $V_0(d)$ as it is again an invariant subspace under all above twirling maps by taking as input the channel difference $I - \mathcal{E}_{\rho^*}$. We will use this fact when giving upper bounds on $\gamma$.

4.2 Proving Measures are Approximate Twirls

In Theorem 8 a key requirement is for the ideal twirling map $\Lambda^*_R$ to be a $\gamma$-approximate twirl with $\gamma < 1$. Here, we give several possible ways to upper bound $\gamma$. A trivial upper bound is 2 which follows directly from Proposition 1.

We establish some notation. Note that a probabilistic distribution $\mu$ on the set $S$ together with the ideal map $\omega : S \to SU(d)$ defines a probabilistic distribution on $SU(d)$, from which the ideal twirling map $\Lambda^*_R$ is uniquely determined. For the rest of the section, we consider probabilistic distributions $\mu$ on $SU(d)$, and let

$$\Lambda(\mu) := \int_{g \sim \mu} \omega(g)^\dagger \circ \cdot \omega(g) dg.$$ 

4.2.1 Bounds on $\gamma$ for Convex Combinations

One straightforward yet somewhat restrictive way of bounding $\gamma$ is when $\mu$ is a convex combination of a unitary 2-design $\nu$ with any other measure.

We first observe that if $\mu$ is a $p$-convex combination of some measure $\mu'$ and $\nu$, then

$$\Lambda(\mu) = \Lambda(p\mu' + (1 - p)\nu) = p\Lambda(\mu') + (1 - p)\Lambda(\nu).$$

Note that this implies the set of twirling maps corresponding to some measure is convex. However, this implies

$$|||\Lambda(\mu) - \Lambda(\nu)|||_\diamond = ||p\Lambda(\mu') + (1 - p)\Lambda(\nu)||_\diamond \leq p|||\Lambda(\mu')|||_\diamond + (1 - p)|||\Lambda(\nu)|||_\diamond = 2p.$$ 

Moreover, $\Lambda(\nu) = \Lambda(\eta)$, where $\eta$ is the Haar random distribution on $SU(d)$. We conclude the following:

**Proposition 10.** If $\mu$ is a $p$-convex combination of any measure and a unitary 2-design, then $\mu$ is a $2p$-approximate twirl.

Consider the Clifford group $C(d) \subseteq SU(d)$ and let the measure $\mu_C$ be the uniform distribution over $C(d)$. Suppose a discrete measure $\mu$ has support that includes $C(d)$. Then, define

$$m := \min_{c \in C(d)} \mu(c) > 0.$$ 

We have that

$$m \times |C(d)| \leq \sum_{c \in C(d)} \mu(c) \leq 1.$$
If $m$ satisfies this with equality then $\mu = \mu_C$. Assume it doesn’t. Then, define the measure $\mu'$ as
\[
\mu' = \frac{1}{1 - m|C(d)|} (\mu - m|C(d)|\mu_C).
\]
By construction, $\mu'$ is a legitimate measure. That is, $\mu$ is a $(1 - m|C(d)|)$-convex combination of $\mu'$ and $\mu_C$. Note that $\mu$ having support that includes $C(d)$ is also a necessary condition for it to be a nontrivial convex combination involving $\mu_C$. By the above result, we can conclude $\mu$ is a $2(1 - m|C(d)|)$-approximate twirl. We can therefore conclude

**Proposition 11.** If a measure $\mu$’s support includes the Clifford group with all probabilities greater than half of that of the uniform distribution over the Clifford group, it is a $\gamma$-approximate twirl with $\gamma < 1$.

In particular, we can find another distribution over the Clifford group for which we can do URB. This can be extended to any unitary 2-design that is a uniform distribution over a finite set. We can also consider more general two-designs. Let $\nu$ be a measure with support a finite subset $S \subseteq SU(d)$ that is a unitary 2-design. We define
\[
M_S(\nu) := \max_{g \in S} \nu(S) > 0.
\]
We have
\[
M_S(\nu) \times |S| \geq \sum_{g \in S} \nu(g) = 1.
\]
Consider a measure $\mu$ whose support includes $S$. Define
\[
m_S(\mu) := \min_{g \in S} \mu(S).
\]
Now,
\[
m_S(\mu) \times |S| \leq \sum_{g \in S} \mu(g) \leq 1.
\]
We conclude $m_S(\mu) \leq M_S(\nu)$. Then, we can define the measure
\[
\mu' := \frac{1}{1 - m_S(\mu)/M_S(\nu)} \left( \mu - \frac{m_S(\mu)}{M_S(\nu)} \nu \right).
\]
By construction, $\mu'$ is a legitimate measure. Thus, $\mu$ is a convex combination of $\mu'$ and $\nu$. Note that $\mu$ having support on $S$ is a necessary condition for $\mu$ to be a nontrivial convex combination involving $\nu$. We conclude

**Proposition 12.** Given a unitary 2-design $\nu$ with finite support $S$, if a measure $\mu$ has probability greater than half of the max probability of $\nu$ on all of $S$, $\mu$ is a $\gamma$-approximate twirl with $\gamma < 1$.

Note that this implies if $M_S(\nu) \geq 2/|S|$,\[
\sum_{g \in S} \mu(g) \geq m_S(\mu) \times |S| > \frac{1}{2} M_S(\nu) \times |S| \geq 1,
\]
which is impossible. Thus in this case we cannot find such a measure $\mu$. Note that we can easily extend these arguments to the case when the support of $\mu$ is not finite.
4.2.2 Bounds on \( \gamma \) via Pauli Representation

For the following bounds we restrict ourselves to the case of qubits, that is \( d = 2^n \). We consider the Pauli matrices \( \{ P_i \}_{i=1}^{d^2} \), that is tensor products of \( \{ I, X, Y, Z \} \), that span \( L(\mathbb{C}^d) \) and are orthogonal under the Hilbert-Schmidt inner product

\[
\text{tr}(P_i P_j) = d \delta_{ij}.
\]

With this basis, we can express any linear operator. Since superoperators can be expressed in terms of left and right multiplications by operators, in general we can express a superoperator as a matrix \( \alpha \) whose coefficients are defined by:

\[
\mathcal{T} = \sum_{i,j} \alpha_{i,j} P_i \cdot P_j.
\]

We refer to this as the \( \alpha \) matrix and will sometimes specify which superoperator we’re considering by \( \alpha(\mathcal{T}) \). Unless otherwise specified, summation indices range from 1 to \( d^2 \). Going further, given a supersuperoperator \( \Lambda \), we now introduce the \( \beta \) tensor whose coefficients are defined by

\[
\Lambda : P_i \cdot P_j \mapsto \sum_{k,l} \beta_{i,j}^{k,l} P_k \cdot P_l.
\]

We can linearly extend this to obtain the action of \( \Lambda \) on any superoperator. We refer to this as the \( \beta \) tensor. The supersuperoperators we will mainly be considering are twirling maps corresponding to a measure \( \mu \), and we sometimes write \( \beta(\mu) \) as the \( \beta \) tensor for \( \Lambda(\mu) \). For a more detailed introduction to these representations and some basic facts that will be relevant to proving upper bounds on \( \gamma \), see Section A.

The first bound is the following:

**Proposition 13.** Let \( \beta(\mu) \) be the \( \beta \) tensor for the twirling map \( \Lambda(\mu) \). Then,

\[
\|\| \Lambda(\mu) - \Lambda(\eta) \|\|_\diamond \leq 2 \left( \sum_{i,j=1}^{d^2} | \beta_{i,j}^{k,l}(\mu) - \beta_{i,j}^{k,l}(\eta) | \right)^{\frac{1}{2}}.
\]

**Proof.** We prove this in Section B.

This is a direct and general upper bound on the diamond norm. However, its form is reminiscent of an \( \ell_2 \)-norm, while we would expect something that looks more like an induced \( \ell_1 \)-norm. We can prove such a bound for a special case.

**Proposition 14.** Suppose

\[
\Lambda(\mu) = \Lambda(\mu) \circ \Lambda(\mu_P),
\]

where \( \mu_P \) is the uniform distribution over Pauli operators. Furthermore, let \( \mathcal{T} \) be a difference of two channels: \( \mathcal{T} = N_1 - N_2 \). Then,

\[
\|\| [\Lambda(\mu) - \Lambda(\eta)](\mathcal{T}) \|\|_\diamond \leq \max_{i \in [d^2]} \sum_{k,l=1}^{d^2} | \beta_{i,i}^{k,l}(\mu) - \beta_{i,i}^{k,l}(\eta) | \|\| \mathcal{T} \|\|_\diamond
\]

where \( [d^2] := \{1, 2, \cdots, d^2\} \).
Proof. We prove this in Section C.

Intuitively, the Pauli representation lets us express supersuperoperators as a linear map on matrices, and our upper bound on the corresponding induced diamond norm resembles the induced ℓ₁-norm for regular matrices:

\[ \|M\|_{1 \rightarrow 1} = \max_j \sum_i |M_{i,j}|. \]

The assumption that \( \Lambda(\mu) \) is right invariant under \( \Lambda(\mu_P) \) is natural in that this allows us to connect the trace norm of \( \alpha \) with the superoperator’s diamond norm. This right invariance is easily enforced by adding a uniformly random Pauli operator before each random gate for any URB scheme, along with the fact that \( \Lambda(\mu_P) \) is a projector as shown in Section C. Finally, the restriction that \( T \) is a difference of channels still allows us to bound \( \gamma \) in Theorem 8. This is because of Remark 9 and the fact that

\[ V_0(d) = \{ c \cdot (N_1 - N_2) | c \in \mathbb{R}, N_i \in \mathcal{C}(d) \}. \]

To see this, we simply observe that a linear combination of a difference of channels is simply a scaled difference of channels. Without loss of generality, let \( c_1, c_2 \geq 0 \) and consider

\[ c_1 (N_1 - N_2) + c_2 (N_3 - N_4) = (c_1 + c_2) \cdot \left[ \left( \frac{c_1}{c_1 + c_2} N_1 + \frac{c_2}{c_1 + c_2} N_3 \right) - \left( \frac{c_1}{c_1 + c_2} N_2 + \frac{c_2}{c_1 + c_2} N_4 \right) \right]. \]

As an exercise, we can evaluate \( \gamma \) for \( \mu_P \):

\[ \max_{i \in [d^2]} \sum_{k,l=1}^{d^2} |\beta_{i,i}^{k,l}(\mu) - \beta_{i,i}^{k,l}(\eta)| = \max_{i > 1} \sum_{k,l=1}^{d^2} |\beta_{i,i}^{k,l}(\mu) - \beta_{i,i}^{k,l}(\eta)| = \left| 1 - \frac{1}{d^2 - 1} \right| + (d^2 - 2) \cdot \left| 0 - \frac{1}{d^2 - 1} \right| = 2 \frac{d^2 - 2}{d^2 - 1}. \]

This for \( d = 2 \) is already \( 4/3 > 1 \), so this does not justify uniform Pauli randomized benchmarking. However, as \( d \rightarrow \infty \), the upper bound goes to \( 2 \), which is the highest the induced diamond norm distance can be.

### 4.3 Approximate Twirls in Other Norms

Theorem \[8\] shows that a single exponential decay can be observed when \( R \) is a twirling scheme with respect to the induced diamond norm. We give upper bounds on the induced diamond norm, but in general this is difficult to compute. Other norms are easier to compute and characterize, and we give the following corollaries as alternatives to Theorem \[8\].

**Corollary 15** (Single exponential decay with respect to the trace norm). Let \( R = (S, \mu, \phi, M, \rho_0) \) be an \((\epsilon, \delta, \gamma)\) twirling scheme with respect to the tuple \((M_0, \phi^*, I, \omega)\) under the trace norm. Given that \( \delta \leq \frac{1 - \gamma}{11} \), there exists \( A, B \in \mathbb{R} \) and \( p \in [1 - 2\delta, 1] \) such that, there exists \( A, B \in \mathbb{R} \) and \( p \in [1 - 2\delta, 1] \) such that

\[ |p_R(m) - (A + Bp^m)| \leq \epsilon + 16(\gamma + 6\delta)^m. \]
Proof. This can be readily proved by replacing the diamond norms in the proof of Theorem 8 by the trace norm.

Given a twirling map, it is often much easier to observe its spectrum through diagonalization of its matrix representation. In contrast, the diamond norm or trace norm are typically computed via semidefinite programming. The following gives an analog of Theorem 8 in terms of a spectral gap, although with an error term involving the dimension of the quantum system.

Corollary 16 (Single exponential decay with respect to the Frobenius norm). Let \( R = (S, \mu, \phi, M, \rho_0) \) be an \((\epsilon, \delta, \gamma)\) twirling scheme on quantum systems with dimension \( d \) with respect to the tuple \((M_0, \phi^*, \mathcal{I}, \omega)\) under the \(\|\cdot\|_2\) norm. Let \( \delta' := d \cdot \delta \). Given that \( \delta' \leq \frac{1-\gamma}{4\gamma} \), there exists \( A, B \in \mathbb{R} \) and \( p \in [1 - 2\delta', 1] \) such that

\[
|p_R(m) - (A + Bp^m)| \leq \epsilon + 16d^{3/2}(\gamma + 6\delta')^m.
\]

Furthermore, in the case that \( \phi^* \) or \( \phi \) maps to unitary mixtures we can restrict to \( \delta' = \sqrt{d} \cdot \delta \).

Proof. There are only two places in the proof of Theorem 8 where the diamond norm is explicitly used, and we replace them with the spectral norms. The first one is at Equation (4.2), to bound the norm of the perturbation \( E \). Here we instead have

\[
\|E\| = \|\Lambda_R - \Lambda_R^*\|_2
\]

\[
= \|\int_{g \sim \mu} dg (\omega(g) \odot \omega(g) - \phi^*(g) \circ \phi(g))\|_2
\]

\[
= \|\int_{g \sim \mu} dg ((\omega(g) \odot \phi^*(g)) \circ \omega(g) + \phi^*(g) \circ (\omega(g) - \phi(g)))\|_2
\]

\[
\leq \int_{g \sim \mu} dg \|\|\omega(g) \odot \phi^*(g)\|_2\|\omega(g)\|_2 + \int_{g \sim \mu} dg \|\phi^*(g) \circ (\omega(g) - \phi(g))\|_2
\]

\[
\leq \sqrt{d} \int_{g \sim \mu} dg \|\omega(g) \odot - \phi^*(g)\|_2 + \int_{g \sim \mu} dg \|\omega(g) - \phi(g)\|_2
\]

\[
\leq d \int_{g \sim \mu} dg \|\omega(g) \odot - \phi^*(g)\|_2 + \int_{g \sim \mu} dg \|\omega(g) - \phi(g)\|_2
\]

\[
\leq d\delta,
\]  

(4.4) where in the third inequality we used the fact that \( \|\omega(g)\|_2 = 1 \leq \sqrt{d} \) and \( \phi^*(g) \) have 2 norm upper bounded by \( \sqrt{d} \). In case that \( \phi^*(g) \) is a unitary mixture (similarly for \( \phi(g) \)), the bound can further be tightened to \( \sqrt{d}\delta \) by observing \( \|\phi^*(g)\|_2 \leq \|\phi^*(g)\|_{tr} = 1 \). The second one is at Equation (4.3). Here we have

\[
\text{tr}[M_0 \cdot L_2(A_2^m R_2(\mathcal{I}) \rho_0)] \leq \|L_2(A_2^m R_2(\mathcal{I}))\|_{tr}
\]

\[
\leq \|L_2(A_2^m R_2)\|_{tr}
\]

\[
\leq d^{3/2} \cdot \|L_2(A_2^m R_2)\|_2
\]

\[
\leq d^{3/2} \cdot \|L_2\|_2 \cdot \|A_2^m\|_2 \cdot \|R_2\|_2
\]

\[
\leq 16d^{3/2} \cdot \|A_2^m\|_2
\]

\[
\leq 16d^{3/2} \cdot (\gamma + 6\delta')^m.
\]  

(4.5)
As in the case of Theorem 8, the crucial property we need to establish is $\gamma < 1$, but now with respect to these other norms.

4.4 Interpretation of the Decay Rate

4.4.1 Gauge Invariance

RB schemes are known to be gauge invariant: that an implementation map $\phi$ only needs to be close to a conjugation of the ideal map $\omega$ in order to exhibit the desired exponential decay, even if $\phi$ and $\omega$ were far apart under diamond norm. This is still true in the URB setting, and gives us the following alternative definition for the near-ideal implementation.

**Definition 17** (Near-ideal implementation under gauge). A URB scheme $R = (S, \mu, \phi, M, \rho_0)$ with $\epsilon$ approximate factoring of $M$ into $(M_0, \phi^*, \mathcal{I})$ is said to have a $(\delta, \kappa)$ near-ideal implementation, if there exists an ideal map $\omega : S \to \mathcal{C}(d)$, and gauges $\mathcal{U}, \mathcal{V}$ being invertible real superoperators, such that

$$\|\|\mathcal{U}\|\|_{\diamond} \|\|\mathcal{U}^{-1}\|\|_{\diamond} \|\|\mathcal{V}\|\|_{\diamond} \|\|\mathcal{V}^{-1}\|\|_{\diamond} \leq \kappa,$$

and

$$E_{g \sim \mu} \left[ \|\mathcal{U} \circ \phi(g) \circ \mathcal{U}^{-1} - \omega(g)\|_{\diamond} + \|\mathcal{V} \circ \phi^*(g) \circ \mathcal{V}^{-1} - \omega(g)^{\dagger}\|_{\diamond} \right] \leq \delta.$$

With the definition of near-ideal implementation under gauge, we can further define that a URB scheme $R$ is a $(\epsilon, \delta, \gamma, \kappa)$ twirling scheme under gauge if it has such a near-ideal implementation.

The gauge free version corresponds to the case that $\mathcal{U} = \mathcal{V} = \text{id}$ and $\kappa = 1$. Our single exponential decay results for different norms can be readily applied to twirling schemes under gauge. We have the following.

**Corollary 18** (Single exponential decay under gauge). Let $R = (S, \mu, \phi, M, \rho_0)$ be an $(\epsilon, \delta, \gamma, \kappa)$ twirling scheme with respect to the tuple $(M_0, \phi^*, \mathcal{I}, \omega, \mathcal{U}, \mathcal{V})$, under the diamond norm or the trace norm. Given that $\delta \leq 1 - \frac{1}{11}$, there exists $A, B \in \mathbb{R}$ and $p \in [1 - 2\delta, 1]$ such that

$$|p_R(m) - (A + Bp^m)| \leq \epsilon + 16\kappa(\gamma + 6\delta)^m.$$

Alternatively, let $R = (S, \mu, \phi, M, \rho_0)$ be an $(\epsilon, \delta, \gamma, \kappa)$ twirling scheme on quantum systems with dimension $d$ with respect to the tuple $(M_0, \phi^*, \mathcal{I}, \omega, \mathcal{U}, \mathcal{V})$ under the Frobenius norm. Let $\delta' := d \cdot \delta$. Given that $\delta' \leq 1 - \frac{1}{11}$, there exists $A, B \in \mathbb{R}$ and $p \in [1 - 2\delta', 1]$ such that

$$|p_R(m) - (A + Bp^m)| \leq \epsilon + 16\kappa d^{3/2}(\gamma + 6\delta')^m.$$

Furthermore, in the case that $\phi^*$ or $\phi$ maps to unitary mixtures we can restrict to $\delta' = \sqrt{d} \cdot \delta$.

**Proof.** The proofs closely mimic the proofs for Theorem 8, Corollary 15 and Corollary 16. The central difference is to substitute the use of $\Lambda_R$ with a gauge-corrected twirling map

$$\tilde{\Lambda}_R := \int_{g \sim \mu} dg \mathcal{V} \circ \phi^*(g) \circ \mathcal{V}^{-1} \circ \mathcal{U} \circ \phi(g) \circ \mathcal{U}^{-1},$$

with the observation that

$$\text{tr}[M_0 \cdot \Lambda_R^0(\mathcal{I})(\rho_0)] = \text{tr}[M_0 \cdot \mathcal{V}^{-1} \circ \tilde{\Lambda}_R^0(\mathcal{V} \circ \mathcal{I} \circ \mathcal{U}^{-1}) \circ \mathcal{U}(\rho_0)].$$
By treating $\tilde{\Lambda}_R$ as a perturbed version of $\Lambda_R^*$, we can similarly get a diagonalization $(L_1, L_2, R_1, R_2, A_1', A_2')$ of $\tilde{\Lambda}_R$. This gives

$$\text{tr}[M_0 \cdot \Lambda_R^m(I)(\rho_0)] = \text{tr}[M_0 \cdot \mathcal{V}^{-1} \circ L_1(A_1')^m R_1(V \circ I \circ U^{-1}) \circ U(\rho_0)] + \text{tr}[M_0 \cdot \mathcal{V}^{-1} \circ L_2(A_2')^m R_2(V \circ I \circ U^{-1}) \circ U(\rho_0)]. \quad (4.6)$$

The second term is an exponential decay term; we have

$$\text{tr}[M_0 \cdot \mathcal{V}^{-1} \circ L_2(A_2')^m R_2(V \circ I \circ U^{-1}) \circ U(\rho_0)] 
\leq ||| \mathcal{V}^{-1} \circ L_2(A_2')^m R_2(V \circ I \circ U^{-1}) \circ U |||_\infty 
\leq ||| \mathcal{V}^{-1} |||_\infty \cdot |||U|||_\infty \cdot |||L_2(A_2')^m R_2|||_\infty \cdot |||V \circ I \circ U^{-1}|||_\infty 
\leq \kappa \cdot |||L_2(A_2')^m R_2|||_\infty,$$

from which Equation (4.3) or Equation (4.5) apply.

The analysis of the first term is similar to that of the proof of Theorem 4.2.

4.4.2 Relating the Decay Rate to the Average Fidelity

The decay rate in an RB or URB scheme is often believed to indicate the average “quality” of a collection of gate implementations. Mathematically, it corresponds to the second largest eigenvalue (guaranteed to be real when the URB scheme is a twirling scheme) of the twirl operator $\Lambda_R$. Relating such an eigenvalue with an indicator of the average “quality” of gates can be non-obvious and sometimes tricky [27]. Of course, we can always avoid this question by defining a URB scheme’s decay rate as a figure of merit by fiat, but we can consider some cases where there is a relatively clear connection between the average fidelity and the second largest eigenvalue.

During the proof of Theorem 4.2 we see that the eigen-channel with eigenvalue 1 is always a replacement channel $\mathcal{E}_{p^*}$. All other eigen-superoperators, including the one corresponding to the second largest eigenvalue, lie in the space of channel differences $V_0(d)$. To see this, note that any eigne-channel $\mathcal{N} \in \mathcal{C}(d)$ of $\Lambda_R$ must be of eigenvalue 1 (otherwise $\Lambda_R(\mathcal{N})$ is no longer a channel) and consequently corresponds to an eigen-superoperator $\mathcal{N} - \mathcal{E}_{p^*} \in V_0(d)$. Let $\mathcal{D}_R \in V_0(d)$ be the eigen-superoperator corresponding to the eigenvalue $p_R = \lambda_2$. An easy case that $p$ has a natural interpretation is when the line $\{\mathcal{E}_{p^*} + \lambda \cdot \mathcal{D}_R | \lambda \in \mathbb{R}\}$ contains a unitary channel $\mathcal{U}$. In this case we can perform a unitary gauge transformation $\phi(g) \mapsto \mathcal{U} \circ \phi(g) \circ \mathcal{U}^{-1}$ so that we can assume without loss of generality that the unitary channel is the identity channel id[6] Then

$$\Lambda_R(\text{id}) = p \cdot \text{id} + (1 - p) \cdot \mathcal{E}_{p^*}.$$

The decay rate $p$ is then related to the fidelity of the channel $\Lambda_R(\text{id}) = \int_{g \sim \mu} d\phi \phi^*(g) \phi(g)$, which in turn is the average fidelity of the channels $\phi^*(g) \phi(g)$ under the probability distribution $\mu$, specifically,

$$p_R = \frac{\int_{g \sim \mu} [F_{\text{avg}}(\phi^*(g) \phi(g))] - 1}{d - 1},$$

6This can be done in general when the line $\{\mathcal{E}_{p^*} + \lambda \cdot \mathcal{D}_R | \lambda \in \mathbb{R}\}$ contains an invertible superoperator $\mathcal{R}$. However in this case the gauge transformation $\phi(g) \mapsto \mathcal{R} \circ \phi(g) \circ \mathcal{R}^{-1}$ may no longer preserve channels, making the physical interpretation of the decay rate less physical.
where the average fidelity of a channel \( C \) is defined as
\[
F_{\text{avg}}(C) := \int_U dU F(U|0\rangle\langle 0|U^\dagger, C(U|0\rangle\langle 0|U^\dagger))
\]
over Haar random unitaries \( U \). One example of an error model satisfying this condition is the \textit{gate-dependent replacement model}: that there exists probabilities \( p(g) \) and states \( \rho(g) \) for each \( g \), such that
\[
\phi^*(g)\phi(g) = p(g) \cdot \text{id} + (1 - p(g)) \cdot \mathcal{E}_{\rho(g)},
\]
\[
\int_{g \in \mu} dg (1 - p(g)) \rho(g) \propto \rho^*.
\]

For gate independent noise models
\[
\phi^*(g) = \mathcal{N}^{(2)}_L \circ \omega(g)^\dagger \circ \mathcal{N}^{(2)}_L, \phi(g) = \mathcal{N}^{(1)}_R \circ \omega(g) \circ \mathcal{N}^{(1)}_R,
\]
we can without loss of generality consider either \textit{in-between noise models}
\[
\phi^*(g) = \omega(g)^\dagger \circ \mathcal{N}_L, \phi(g) = \mathcal{N}_R \circ \omega(g)
\]
or \textit{sandwiched noise models}
\[
\phi^*(g) = \mathcal{N}_L \circ \omega(g)^\dagger, \phi(g) = \omega(g) \circ \mathcal{N}_R
\]
through gauge transformation, where \( \mathcal{N}_L := \mathcal{N}^{(2)}_L \circ \mathcal{N}^{(1)}_L \) and \( \mathcal{N}_R := \mathcal{N}^{(2)}_R \circ \mathcal{N}^{(1)}_R \).

- For the \textit{in-between noise model}, we have
\[
\Lambda_R(\text{id}) = \Lambda^*_R(\mathcal{N}_L \circ \mathcal{N}_R).
\]
In the case that \( \mu \) gives rise to a unitary 2-design, or equivalently \( \Lambda^*_R = \Lambda^* \), we have
\[
\Lambda_R(\text{id}) = \Lambda^*_R(\mathcal{N}_L \circ \mathcal{N}_R) = p \cdot \text{id} + (1 - p) \cdot \text{dep},
\]
where \( p_R = \frac{d - F_{\text{avg}}(\mathcal{N}_L \circ \mathcal{N}_R)^{-1}}{d - 1} \). This recovers the case that RB on unitary two-designs extract the average fidelity of gate-independent noises.

- For the \textit{sandwiched noise model}, we have
\[
\Lambda_R(\text{id}) = \mathcal{N}_L \circ \mathcal{N}_R.
\]
Then the RB exponent \( p_R \) relates to the average fidelity of \( \mathcal{N}_L \circ \mathcal{N}_R \) if the following holds:
\[
\exists p \in [0, 1], \mathcal{N}_L \circ \mathcal{N}_R = p \cdot \text{id} + (1 - p) \cdot \mathcal{E}_{\rho^*},
\]
where \( \rho^* \) is the eigenstate of the channel \( \int_{g \sim \mu} dg \phi^*(g) = \mathcal{N}_L \circ \int_{g \sim \mu} dg \omega(g)^\dagger \). This also guarantees that \( p_R = p \).

To summarize, for gate-independent noise models, the RB exponent extracted from the experiment is not always determined by the average fidelity of the error channels, partially because that the ideal twirl is not always a full twirl. Special cases where there is such a determination is when the ideal twirl is a full twirl, and when the noise channels adds up to an replacement channel with the maximal eigenstate.
4.5 Analysis on Robustness of Data Fitting

In this section, we will prove that a sufficiently small perturbation to an exponential decay curve will not significantly affect the extracted decay rate. More specifically, we will prove the following lemma:

Lemma 19. Let \( C_0(m) = A_0 \alpha^m + B_0 \) be an exponential decay curve. Given \( \epsilon \ll A_0 \alpha^M / 2 \), if there is another exponential curve \( C_1(m) = A_1 \beta^m + B_1 \) such that \( |C_1(m) - C_0(m)| \leq \epsilon \) for all integers \( m \geq M \), then the decay rates \( \beta \) and \( \alpha \) must satisfy

\[
|\beta - \alpha| \leq 2\epsilon \frac{\alpha + 1}{A_0 \alpha^M - 2\epsilon}.
\]

Proof. Since \( |B_0 - B_1| = \lim_{m \to \infty} |C_0(m) - C_1(m)| \leq \epsilon \), we have \( |A_1 \beta^m - A_0 \alpha^m| \leq 2\epsilon \) for any integer \( m \geq M \).

Assume without loss of generality that \( 1 > \beta \geq \alpha > 0 \). Then we have

\[
\beta (A_0 \alpha^M - 2\epsilon) \leq \beta A_1 \beta^M = A_1 \beta^{M+1} \leq A_0 \alpha^{M+1} + 2\epsilon
\]

\[
\Rightarrow |\beta - \alpha| \leq \frac{2\epsilon A_0 \alpha^{M+1} + 2\epsilon}{A_0 \alpha^M - 2\epsilon}.
\]

(4.7)

Lemma 19 bounds the deviation of the decay rate, given that the fitted curve is \( O(\epsilon/A_0) \)-close to a certain ground truth single exponential decay curve under the \( \ell_\infty \)-norm. In our case, such \( \ell_\infty \) deviations come from the error terms \( \epsilon + 16(\gamma + 6\delta)^m \), standard deviation from sampling \( \Theta \left( \log k / \sqrt{K} \right) \) (\( k \) being the number of sequence lengths chosen and \( K \) the number of repeats for each sequence length), and imperfect fitting algorithms. Even though finite \( \epsilon \) makes the error term non-vanishing, we conclude that unless the number of samples for each sequence length exceeds \( \epsilon^{-2} \), the inaccuracy of decay rate extraction mainly comes from stochastic fluctuations of the random experiments.

5 Discussion and Open Questions

In this paper we propose a new framework for randomized benchmarking, which we call the URB framework. To formulate this we take major components of widely used benchmarking schemes such as group-based RB and linear XEB and then generalize them. With this generalization we can express a much wider variety of schemes that could possibly overcome the shortcomings of current schemes, such as scalability. For a certain class of URB schemes, which we call twirling schemes, we can prove an exponential decay using Theorem 8, a hallmark feature of randomized benchmarking schemes. We saw that the crucial property we need to establish to apply Theorem 8 is for the twirling operator \( \Lambda^*_R \) to be a \( \gamma \)-approximate twirl for \( \gamma < 1 \). We gave upper bounds for \( \gamma \) and give alternative exponential decay results for easier bounds to compute, such as the \( ||| \cdot |||_2 \) norm. We also discuss issues of gauge invariance and the interpretation of the decay rate.

One recent RB variant, called randomized benchmarking with mirror circuits [18], aims to extract system fidelity information from shallow Clifford circuits. The random gate sequence consists of

1. Clifford layers with a mirror structure, i.e. the second half of the Clifford circuits are inversions of the first half.
2. Uniform Pauli gates interleaving the Clifford layers

3. Single-qubit Cliffords at the beginning and the end of the circuits.

This falls into our RB framework as all gates after the first half of Clifford layers can be seen as the post-processing POVM. However, with the presence of the interleaving Pauli gates, it is impossible to separate out “inverting maps” that locally approximately inverts the implementation maps. Hence, this scheme does not readily yield to our analysis. In this scheme, the Pauli error is corrected on a global scale after the final measurement. We leave it to future work to analyze such schemes involving complex correlations between the random gates.

We leave some open questions for future work. Obviously, the value of our framework is really proved by novel benchmarking schemes that have better performance by bypassing the group-based requirement. One work towards this direction is [28], which proposes performing linear XEB with shallow Clifford circuits so that the classical simulation step is much more scalable to larger numbers of qubits. They simulate systems with more than a thousand qubits to showcase this scalability. An earlier work [22] also considered a variant of RB that does not form a group.

The upper bounds on $\gamma$ are somewhat loose and could be analyzed more systematically. In particular, it would be interesting to express the induced diamond norm as an SDP so that it can be bounded or even exactly calculated. Furthermore, various inequalities of the proof of Theorem 5 could be tightened, such as improving the factor of 11 in the $\delta \leq (1 - \gamma)/11$ bound and a possible typicality argument for the bounding of $\|((A_2')^m)\|$. Also, our result on robustness of data fitting naturally leads to a question of sample complexity, which we also leave for future work. As discussed above, the interpretation of the decay rate is still a matter of discussion. Also, the exponential decay of the $F(m)$ quantity for linear XEB defined in Equation (3.1) is still to be further investigated.

Lastly, from our work we recognize the value of a systematic formulation of norms for multilinear algebra as applied to quantum information science.

Another direction to pursue is to recover matrix exponential decay results for group-based RB [4, 6, 5, 21]. There are two major differences between our result and Fourier-analysis based results for group-based RB. First, we treat the final recovery gate as an imperfect implementation of the perfect recovery gate, which enables factorization and hence the study of twirling maps, but this introduces a constant error term which is not present in the Fourier-based analyses. Intuitively, this is because our twirling map approach fails to capture the correlation between the noise in the final recovery gate with those of the previously applied gates due to taking the expectation over the group. This removes the constant error term in Fourier analysis-based approaches. We leave the incorporation of such correlated noise in a twirling maps approach to future work. Also, the case where multiple irreducible representations are concerned corresponds to the case where $\Lambda^*_R \neq \Lambda^*$ is a projector on a larger space and is thus not an approximate twirl. Recovering the matrix exponential result then requires an investigation into $\Lambda^*_R$ that do not form approximate twirls. However, our framework is not concerned with group structure or irreducible representations, but only the twirling map. This could provide a conceptually easier approach when dealing with infinite groups.

**Acknowledgements**

We would like to thank Chunqing Deng, Linghang Kong, Yaoyun Shi, and Tenghui Wang for helpful comments. DD would like thank God for all of His provisions.
A Properties of the $\alpha$ Matrix and $\beta$ Tensor

We first recall that we have a basis $\{P_i\}_{i=1}^{d^2}$ of $L(\mathbb{C}^d)$ that satisfies

$$\text{tr}[P_i P_j] = d\delta_{ij}.$$ 

We can use this basis to express superoperators

$$\mathcal{T} = \sum_{i,j} \alpha_{i,j} P_i \cdot P_j$$

and supersuperoperators

$$\Lambda : P_i \cdot P_j \mapsto \sum_{k,l} \beta_{i,j}^{k,l} P_k \cdot P_l.$$ 

We establish some facts relating the norms of the $\alpha$ matrix with the norms of its corresponding superoperator. We will use these results when proving the bounds on $\gamma$.

**Proposition 20.** Let $\mathcal{T}$ be a superoperator. Then,

$$\|\mathcal{T}\|_{SO} = d\|\alpha(\mathcal{T})\|_{HS},$$

where $\| \cdot \|_{SO}$ is the norm on superoperators induced by the $(\cdot, \cdot)_{SO}$ inner product.

**Proof.** Since the Pauli operators $\{P_m\}_{m=1}^{d^2}$ are orthogonal under the Hilbert Schmidt inner product, we can directly compute

$$\|\mathcal{T}\|_{SO}^2 = \frac{1}{d} \sum_{m=1}^{d^2} \langle \mathcal{T}(P_m), \mathcal{T}(P_m) \rangle_{HS}$$

$$= \frac{1}{d} \sum_{m=1}^{d^2} \text{tr} \left[ \sum_{i,j,k,l=1}^{d^2} \alpha_{i,j}^* \alpha_{k,l} P_j P_m P_k P_m P_l \right]$$

$$= \frac{1}{d} \text{tr} \left[ \sum_{i,j,k,l=1}^{d^2} \alpha_{i,j}^* \alpha_{k,l} P_j (d \text{tr}[P_k] I) P_l \right]$$

$$= \frac{1}{d} \text{tr} \left[ \sum_{i,j,k,l=1}^{d^2} \alpha_{i,j}^* \alpha_{k,l} \delta_{ik} P_j P_l \right]$$

$$= \frac{d^2}{d} \sum_{i,j,k,l=1}^{d^2} \alpha_{i,j}^* \alpha_{k,l} \delta_{ik} \delta_{jl}$$

$$= d^2 \sum_{i,j=1}^{d^2} |\alpha_{i,j}|^2.$$
In the fourth equality we used for any operator $X$,

$$
\sum_{m=1}^{d^2} P_m X P_m = d \text{tr}[X]I.
$$

This directly gives our conclusion.

**Proposition 21.** Let $\mathcal{N}$ be a quantum channel. Its $\alpha$ matrix is positive semidefinite with unit trace.

**Proof.** Let $\mathcal{N}$ be a quantum channel. Then, it has a Kraus representation

$$
\mathcal{N} = \sum_{k=1}^{K} A_k \cdot A_k^\dagger,
$$

where $K \leq d^2$ is the Kraus rank and

$$
\sum_{k=1}^{K} A_k^\dagger A_k = I.
$$

Suppose the decomposition of $A_k \in L(\mathbb{C}^d)$ in terms of $\{P_i\}_i$ is given by

$$
A_k = \sum_i a_{k,i} P_i.
$$

Then,

$$
\alpha(\mathcal{N})_{i,j} = \sum_{k=1}^{K} a_{k,i} a_{k,j}^*.
$$

This implies $\alpha$ is a Gram matrix, which is equivalent to being positive semidefinite. Furthermore,

$$
\sum_{k=1}^{K} A_k^\dagger A_k = \sum_{i,j} \alpha_{j,i} P_i P_j = I.
$$

Taking traces on both sides,

$$
\sum_{i,j} \alpha_{j,i} d \delta_{ij} = d.
$$

Hence, $\alpha$ has unit trace.
For small positive $\epsilon$, $\alpha$ is positive semidefinite with unit trace. However,
\[ \sum_{i,j} \alpha_{j,i} B_i B_j = I + 2\epsilon X \neq I. \]

We next consider the $\beta$ tensor for twirling maps corresponding to a measure. We write out:
\[ \Lambda(\mu) : P_i \cdot P_j \mapsto \int_{g \sim \mu} d g U_g^\dagger P_i U_g \cdot U_g^\dagger P_j U_g. \]
Denote
\[ U_g^\dagger P_i U_g = \sum_{k=1}^{d^2} u(g)_{i,k} P_k. \]
Since the Paulis are Hermitian and orthogonal, $u(g)_{i,k} \in \mathbb{R}$ and $\sum_{k=1}^{d^2} |u(g)_{i,k}|^2 = 1$. Then,
\[ \Lambda(\mu) : P_i \cdot P_j \mapsto \int_{g \sim \mu} d g U_g^\dagger P_i U_g \cdot U_g^\dagger P_j U_g = \int_{g \sim \mu} d g \sum_{k,l=1}^{d^2} u(g)_{i,k} u(g)_{j,l} P_k \cdot P_l, \]
so
\[ \beta_{i,j}^{k,l}(\mu) = \int_{g \sim \mu} d g u(g)_{i,k} u(g)_{j,l} = \frac{1}{d^2} \int_{g \sim \mu} d g \text{tr}[U_g^\dagger P_i U_g P_k] \text{tr}[U_g^\dagger P_j U_g P_l]. \]
This proves $\beta_{i,j}^{k,l}(\mu) \in \mathbb{R}$. From this representation we also obtain the identity:
\[ \beta_{i,j}^{k,l}(\mu) = \beta_{j,i}^{l,k}(\mu) \]
as well as the fact
\[ \Lambda(\mu)(I \cdot I) = I \cdot I. \quad (A.1) \]
Furthermore, since $\Lambda(\mu)$ preserves the set of channels, if $\mathcal{N}$ is a channel,
\[ \Lambda(\mu)(\mathcal{N}) = \sum_{i,j=1}^{d^2} \alpha_{i,j} \sum_{k,l=1}^{d^2} \beta_{i,j}^{k,l} P_k \cdot P_l = \sum_{k,l=1}^{d^2} \left( \sum_{i,j=1}^{d^2} \alpha_{i,j} \beta_{i,j}^{k,l}(\mu) \right) P_k \cdot P_l \]
is also a channel, which puts additional constraints on $\beta(\mu)$.

Lastly, we compute the $\beta$ tensor for the Haar measure: $\beta(\eta)$. We can use the expression for a Haar twirl from \[1\] to compute
\[ \Lambda(\mu)(P_i \cdot P_j)(\rho) = \frac{\text{tr}[P_i P_j] \text{tr}[\rho] I}{d} + \frac{d \text{tr}[P_i] \text{tr}[P_j] - \text{tr}[P_i P_j]}{d(d^2 - 1)} \left( \rho - \frac{\text{tr}[\rho] I}{d} \right). \]
If exactly one of $P_i, P_j$ is $I$, then the RHS is zero. If they’re both not $I$,
\[ \Lambda(\mu)(P_i \cdot P_j)(\rho) = \delta_{ij} \frac{\text{tr}[\rho] I}{d} - \frac{\delta_{ij}}{d^2 - 1} \left( \rho - \frac{\text{tr}[\rho] I}{d} \right) \]
\[ = \frac{\delta_{ij}}{d^2 - 1} \left[ -\rho + d^2 \frac{\text{tr}[\rho] I}{d} \right] \]
\[ = \frac{\delta_{ij}}{d^2 - 1} \left[ -I \rho + \sum_{k=1}^{d^2} P_k \rho P_k \right] \]
\[ = \frac{\delta_{ij}}{d^2 - 1} \sum_{k=2}^{d^2} P_k \rho P_k. \]
Along with Equation (A.1), we conclude
\[
\beta_{i,j}(\eta) = \begin{cases} 
1 & i = j = k = l = 1 \\
\frac{1}{d^2-1} & (i = j > 1) \land (k = l > 1) \\
0 & \text{otherwise}
\end{cases} 
\]  
(A.2)

In words, \( \Lambda(\eta) \) maps a pair of the same non-identity Pauli’s to a uniform convex combination of same non-identity Pauli’s.

**B \ \ell_2\text{-Norm Like Bound}**

Here we prove Proposition 13.

*Proof.* Let \( T \) be a superoperator. We compute
\[
\|\{(\Lambda(\mu) - \Lambda(\mu H))(T)\|_0 \\
\leq \max_{\|\rho\|_1 = 1} \|\{(\Lambda(\mu)) T \otimes I - \Lambda(\mu H))(T) \otimes I(\rho)\|_1 \\
\leq \frac{1}{d} \|\|T\|\|_{SO} \|\{(\Lambda(\mu)) T \otimes I - \Lambda(\mu H))(T) \otimes I(\rho)\|_1 \\
\leq 2\|\|T\|\|_0 \|\{(\Lambda(\mu)) T \otimes I - \Lambda(\mu H))(T) \otimes I(\rho)\|_1 \\
\leq \|\|T\|\|_0 \|\{(\Lambda(\mu)) T \otimes I - \Lambda(\mu H))(T) \otimes I(\rho)\|_1 \\
\]
where the second inequality involves the following argument:
\[
\sum_m |\langle a, b_m \rangle| \leq \max_{k_m \in U(1)} \langle a, \sum_m k_m \cdot b_m \rangle \\
\leq \|a\|_{\ell_2} \cdot \max_{k_m \in U(1)} \|\sum_m k_m \cdot b_m\|_{\ell_2} \\
= \|a\|_{\ell_2} \cdot \max_{k_m \in U(1)} \sqrt{\sum_{m, m'} (k_m \cdot k_{m'}) \langle b_m, b_{m'} \rangle} \\
\leq \|a\|_{\ell_2} \cdot \sqrt{\sum_{m, m'} |\langle b_m, b_{m'} \rangle|},
\]
the third equality follows from Proposition 20 and the last inequality follows from Equation (2.3).
C Induced $\ell_1$-Norm Like Bound

Here we prove Proposition 14:

**Proof.** We first establish that

$$\|\Lambda(\mu)(T)\|_{\diamond} = \left\| \int_{g \sim \mu} \omega(g^{-1}) \circ T \circ \omega(g) dg \right\|_{\diamond} \leq \int_{g \sim \mu} dg \|\omega(g^{-1})\|_{\diamond} \|T\|_{\diamond} \|\omega(g)\|_{\diamond} = \|T\|_{\diamond}. $$

Furthermore, we have that

$$\Lambda(\mu_P)(N_i) = \Lambda(\mu_P) \left( \sum_{i,j} \alpha_{i,j} P_i \cdot P_j \right)$$

$$= \frac{1}{d^2} \sum_{k} \sum_{i,j} \alpha_{i,j} P_k P_i P_k \cdot P_k P_j P_k$$

$$= \frac{1}{d^2} \sum_{k} \sum_{i,j} \alpha_{i,j} (-1)^{(b_k \cdot b_i)_S + (b_k \cdot b_j)_S} P_i \cdot P_j$$

$$= \frac{1}{d^2} \sum_{k} \sum_{i,j} \alpha_{i,j} (-1)^{(b_k \cdot b_i \oplus b_j)_S} P_i \cdot P_j$$

$$= \sum_i \alpha_{i,i} (T) P_i \cdot P_i,$$

(C.1)

where $\langle \cdot, \cdot \rangle_S$ denotes the symplectic inner product of the binary symplectic representations of a Pauli string [29] and $\oplus$ denotes bitwise addition. Two Paulis commute iff the inner product vanishes. The fourth equality follows from the distributive property of the inner product. The last inequality follows because when $i = j$, $b_i \oplus b_j = 0$, while when $i \neq j$, half of the summed $b_k$ is orthogonal to $b_i \oplus b_j$ while the other half is not and therefore cancels out.

Since $N_i$ is a channel, $\alpha(N_i)$ has unit trace. Thus, by Equation (C.1), $\Lambda(\mu_P)(N_i)$ is a Pauli channel, that is, a mixture of Pauli unitaries, and $\Lambda(\mu_P)(T)$ is a difference of Pauli channels. Hence we conclude

$$\|T\|_{\diamond} \geq \|\Lambda(\mu_P)(T)\|_{\diamond}$$

$$= \|\text{diag}(\alpha(T))\|_{\ell_1},$$

(C.2)

where the equality is proven in [30].

By Equation (A.2),

$$\Lambda(\eta) = \Lambda(\eta) \circ \Lambda(\mu_P).$$
We can therefore argue

\[
\| [\Lambda(\mu) - \Lambda(\eta)](\mathcal{T}) \|_{\diamond}
= \| [\Lambda(\mu) - \Lambda(\eta)] \circ \Lambda(\mu P)(\mathcal{T}) \|_{\diamond}

= \max_{\|\rho\|_1 = 1} \| [\Lambda(\mu)(\Lambda(\mu P)(\mathcal{T})) \otimes I - \Lambda(\eta)(\Lambda(\mu P)(\mathcal{T})) \otimes I](\rho) \|_1

= \max_{\|\rho\|_1 = 1} \left\| \sum_{k,l=1}^{d^2} \sum_{i=1}^{d^2} \alpha_{i,i}(\mathcal{T}) \left( \beta^{k,l}_{i,i}(\mu) - \beta^{k,l}_{i,i}(\eta) \right)(P_k \otimes I)\rho(P_l \otimes I) \right\|_1

\leq \sum_{k,l=1}^{d^2} \sum_{i=1}^{d^2} \alpha_{i,i}(\mathcal{T}) \left( \beta^{k,l}_{i,i}(\mu) - \beta^{k,l}_{i,i}(\eta) \right) \left( \max_{\|\rho\|_1 = 1} \| (P_k \otimes I)\rho(P_l \otimes I) \|_1 \right)

\leq \sum_{i=1}^{d^2} \left| \alpha_{i,i}(\mathcal{T}) \right| \sum_{k,l=1}^{d^2} \left| \beta^{k,l}_{i,i}(\mu) - \beta^{k,l}_{i,i}(\eta) \right| \leq \max_{i \in [d^2]} \sum_{k,l=1}^{d^2} \left| \beta^{k,l}_{i,i}(\mu) - \beta^{k,l}_{i,i}(\eta) \right| \| \text{diag}(\alpha(\mathcal{T})) \|_{\ell_1}

\leq \max_{i \in [d^2]} \sum_{k,l=1}^{d^2} \left| \beta^{k,l}_{i,i}(\mu) - \beta^{k,l}_{i,i}(\eta) \right| \| \mathcal{T} \|_{\diamond},

\]

where the last inequality uses Equation (C.2).

References

[1] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347, 2005.

[2] Emanuel Knill, Dietrich Leibfried, Rolf Reichle, Joe Britton, R Brad Blakestad, John D Jost, Chris Langer, Roee Ozeri, Signe Seidelin, and David J Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, 2008.

[3] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Physical review letters*, 106(18):180504, 2011.

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

[5] Seth T Merkel, Emily J Pritchett, and Bryan H Fong. Randomized benchmarking as convolution: Fourier analysis of gate dependent errors. *Quantum*, 5:581, 2021.

[6] Jonas Helsen, Ingo Roth, Emilio Onorati, Albert H Werner, and Jens Eisert. A general framework for randomized benchmarking. *arXiv preprint arXiv:2010.07974*, 2020.

[7] John P Gaebler, Adam M Meier, Ting Rei Tan, Ryan Bowler, Yiheng Lin, David Hanneke, John D Jost, JP Home, Emanuel Knill, Dietrich Leibfried, et al. Randomized benchmarking of multiqubit gates. *Physical review letters*, 108(26):260503, 2012.
[8] Jay M Gambetta, Antonio D Córcoles, Seth T Merkel, Blake R Johnson, John A Smolin, Jerry M Chow, Colm A Ryan, Chad Rigetti, Stefano Poletto, Thomas A Ohki, et al. Characterization of addressability by simultaneous randomized benchmarking. *Physical review letters*, 109(24):240504, 2012.

[9] David C McKay, Sarah Sheldon, John A Smolin, Jerry M Chow, and Jay M Gambetta. Three-qubit randomized benchmarking. *Physical review letters*, 122(20):200502, 2019.

[10] Shely Garion, Naoki Kanazawa, Haggai Landa, David C McKay, Sarah Sheldon, Andrew W Cross, and Christopher J Wood. Experimental implementation of non-clifford interleaved randomized benchmarking with a controlled-s gate. *Physical Review Research*, 3(1):013204, 2021.

[11] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. Characterizing universal gate sets via dihedral benchmarking. *Physical Review A*, 92(6):060302, 2015.

[12] Andrew W Cross, Easwar Magesan, Lev S Bishop, John A Smolin, and Jay M Gambetta. Scalable randomised benchmarking of non-clifford gates. *npj Quantum Information*, 2(1):1–5, 2016.

[13] Joel J Wallman, Chris Granade, Robin Harper, and Steven T Flammia. Estimating the coherence of noise. *New Journal of Physics*, 17(11):113020, 2015.

[14] Joel J Wallman, Marie Barnhill, and Joseph Emerson. Robust characterization of leakage errors. *New Journal of Physics*, 18(4):043021, 2016.

[15] Jun Zhang, Jiri Vala, Shankar Sastry, and K Birgitta Whaley. Geometric theory of nonlocal two-qubit operations. *Physical Review A*, 67(4):042313, 2003.

[16] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[17] Jiaqing Jiang, Xiaoming Sun, Shang-Hua Teng, Bujiao Wu, Kewen Wu, and Jialin Zhang. Optimal space-depth trade-off of cnot circuits in quantum logic synthesis. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 213–229. SIAM, 2020.

[18] Timothy Proctor, Stefan Seritan, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Scalable randomized benchmarking of quantum computers using mirror circuits. *arXiv preprint arXiv:2112.09853*, 2021.

[19] Timothy J Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Direct randomized benchmarking for multiqubit devices. *Physical review letters*, 123(3):030503, 2019.

[20] Alexander Erhard, Joel J Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban A Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. Characterizing large-scale quantum computers via cycle benchmarking. *Nature communications*, 10(1):1–7, 2019.
[21] Linghang Kong. A framework for randomized benchmarking over compact groups. *arXiv preprint arXiv:2111.10357*, 2021.

[22] Kristine Boone, Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. Randomized benchmarking under different gate sets. *Physical Review A*, 99(3):032329, 2019.

[23] Richard A Low. Pseudo-randomness and learning in quantum computation. *arXiv preprint arXiv:1006.5227*, 2010.

[24] GW Stewart and Ji-Guang Sun. *Matrix perturbation theory*. Boston: Academic Press, 1990.

[25] Jonas Helsen, Xiao Xue, Lieven MK Vandersypen, and Stephanie Wehner. A new class of efficient randomized benchmarking protocols. *npj Quantum Information*, 5(1):1–9, 2019.

[26] Daniel Stilck França and AK Hashagen. Approximate randomized benchmarking for finite groups. *Journal of Physics A: Mathematical and Theoretical*, 51(39):395302, 2018.

[27] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout. What randomized benchmarking actually measures. *Physical review letters*, 119(13):130502, 2017.

[28] Jianxin Chen, Dawei Ding, Cupjin Huang, and Linghang Kong. Linear cross entropy benchmarking with clifford circuits. *arXiv preprint arXiv:2206.08293*, 2022.

[29] Daniel Gottesman. *Stabilizer codes and quantum error correction*. PhD thesis, Caltech, 1997. eprint: quant-ph/9705052.

[30] Massimiliano F Sacchi. Optimal discrimination of quantum operations. *Physical Review A*, 71(6):062340, 2005.