Concepts and methods for benchmarking quantum computers

by

Kristine Boone

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Physics

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Robin Blume-Kohout Scientist, Sandia National Laboratories
Supervisor(s):	Joseph Emerson Professor, Applied Mathematics, University of Waterloo
Internal Member:	Raymond Laflamme Professor, Physics and Astronomy, University of Waterloo
Internal-External Member:	Joel Wallman Assistant Professor, Applied Mathematics, University of Waterloo
Other Member(s):	Daniel Gottesman Adjunct Professor, Physics and Astronomy, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

My contributions to all of the works presented in this thesis are major and were all under the supervision of Joseph Emerson and Joel Wallman.

The work in chapter 3 has not been published because it is a relatively small numerical calculation which we hope to publish in combination with other complimentary results in the future. I was the primary contributor of this work.

Chapter 4 is largely from [31], with adjustments to notation for consistency within this thesis. While I was not first author of this work, I made substantial contributions to the results, including performing all of the numerical computations, contributing to the development of the theoretical proofs, and writing substantial portions of the publication.

Chapter 5 is largely from [23], with slight changes to notation for consistency, as well as the addition of figure fig. 5.1. I was the main contributor to this project.

This work from chapter 6 will be published in the coming months, for which I am the primary contributor. I played an important role in choosing the suite of benchmarking protocols, I worked closely with the two collaborators (Honeywell and AQT) on the best way to implement the BGSA, I personally ran the BGSA on numerous IBM devices using IBM's remote access, and did the data analysis for the three devices with support from the TrueQ software.

Abstract

Quantum computers have the potential to bring about a new age of technology, but in order for them to become a reality, the elephant in the room must be addressed, quantum errors. The complexity of quantum errors grows exponentially as the system size is increased; scalable characterization protocols which work under physically realistic noise are therefore needed to assess and optimize control.

Randomized Benchmarking (RB) is a family of characterization protocols with the requisite qualities. RB protocols are independent of state preparation and measurement (SPAM) errors, and because they extract partial information about the error profiles in the form of figures of merit, they are scalable in the number of qubits. In this thesis, we study these figures of merit under various error models, gate-sets, and sample distributions, to further our understanding of their physical meaning, and to fine-tune the best practice for diagnosing errors in basic gate-sets.

The first study presented in this thesis looks at the physical meaning of the unitarity under physically realistic gate-dependent noise. Numerically, we demonstrate that under these realistic noise models, the unitarity obtained via Extended RB (XRB) overlaps strongly with the average of the unitarity of the errors over the set of individual operations.

The second study numerically and analytically proves that the Standard RB (SRB) decay parameter p coincides with the decay parameter of the novel figure of merit *gate-set circuit fidelity*. We also prove that, in the limit of high-fidelity single-qubit experiments, the disconnect between the average gate fidelity and SRB experimental results is caused by a basis mismatch between the gates and the state-preparation and measurement procedures.

In the third part of this thesis, we compare the results of two variations of the SRB protocol, one that samples uniformly (SRB) and one that samples non-uniformly (NIST RB) from the Clifford group. We prove that NIST RB produces an exponential fidelity curve that gives different infidelity results than SRB.

The fourth and final study gives a standard for assessing basic gate-sets. Experimental results are given from the assessment run on super-conducting and ion-trap devices. We also experimentally show how the systematic error bounds can be tightened by accounting for the error's degree of coherence, as well as using weaker randomizing sets.

Acknowledgements

I would like to thank Joseph Emerson, Joel Wallman, and Arnaud Carignan-Dugas for their invaluable guidance. A special thank you to Stefanie Beale for all your feedback in the writing of this thesis. I would also like to thank my family, the Muffins, and the entire IQC community; you kept me sane and motivated over the last few years, and for that I will be forever grateful.

Last, but certainly not least, I want to thank Doug Braden for the endless love and support.

Dedication

I want to dedicate this thesis to every girl who has been told *"girls aren't good at math"*.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

RB Randomized Benchmarking v

SPAM State Preparation And Measurement v, 2, 24–26

XRB Extended Randomized Benchmarking from protocol 3 v, 27, 33

SRB Standard Randomized Benchmarking from protocol 1 v, 26, 31

NIST National Institute of Standards and Technology v

NIST RB Randomized Benchmarking protocol from protocol 6 v

AQT Advanced Quantum Testbed xiv, 87, 97

LBNL Lawrence Berkeley National Laboratory xiv, 87, 97

NISQ Noisy Intermediate-Scale Quantum Computing 1, 85

POVM Positive Operator-Value Measure 8

PVM Projector Valued Measure 8

CPTP Completely Positive Trace Preserving 8

QCVV Quantum Characterization, Verification, & Validation 24

QST Quantum State Tomography 24

QPT Quantum Process Tomography 25

GST Gate Set Tomography 26

CB Cycle Benchmarking from protocol 4 27, 34

KNR K-body Noise Reconstruction 27

IRB Interleaved Randomized Benchmarking from protocol 2 32

FTQC Fault-Tolerant Quantum Computation 85

BGSA Basic Gate-Set Assessment 85

List of Symbols

$M_d(\mathbb{C})$ d x d matricies with complex entries 4

Tr Trace operator, for a matrix $A \in M_d(\mathbb{C})$, $\mathrm{Tr}\{A\} = \sum_i [A]_{ii}$ 4

$\mathbb{I}$ Identity Operator 8, 9

X Pauli matrix, σ_X 9

Y Pauli matrix, σ_Y 9

Z Pauli matrix, σ_Z 9

$\mathbb{P}_n$ The n-qubit Pauli group. 11

$\mathbb{C}_n$ The n-qubit Clifford group. 11

SU(d) The n-qubit Special Unitary Group for d=2^n. 11

$\mathbb{U}_n$ The Unitary Group is the set of all $d \times d$ unitary matricies for which $U^\dagger U = \mathbb{I}$ 11

$\tilde{\mathcal{G}}$ Noisy implementation of ideal gate ($\mathcal{G}$) 16

$\mathcal{G}$ Ideal Gate 16

$\mathbb{E}$ Expectation value of random variable. 17

F Average gate fidelity 17

r Infidelity 17

p Effective depolarizing parameter 17, 26, 47, 48

χ_{00} Process Fidelity 17

e_F Process Infidelity 17

u Unitarity 17

$\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m)$ Gate-set circuit fidelity 47

$\mathbb{N}$ The single-qubit NIST gate-set used in NIST RB. 72

e_S Stochastic process infidelity 87

Chapter 1

Introduction

Large scale quantum computers have the potential to revolutionize our world through material design[13], quantum chemistry[110], drug discovery[29], quantum encryption[54], and the boundless applications that have yet to be discovered. The Achilles heel to its development is that quantum error, ubiquitously present in quantum systems, becomes exponentially more complex as system size is increased.

Thankfully, assuming that the error does not exceed some threshold [3, 91, 93, 4, 130, 91], the error can be corrected. Therefore, we need scalable characterization techniques to assess the errors in the elementary gate-set with enough detail to infer the quality of the overall computation.

This thesis explores the methods and assumptions of various error estimation protocols in order to improve pulse-engineering approaches and create more reliable, robust, and potentially automated calibration methods.

As an explosively growing field, there is an ever increasing demand for a widely accepted benchmarking standard in order to bring continuity to the chaos. By furthering our understanding of numerous protocols and figures of merit, we have been able to create this widely accepted standard which obtains the most relevant information about the single- and 2-qubit error profile, while limiting the overhead.

To ensure computational advantage, a characterization method must require fewer classical computational resources than simply simulating the ideal quantum circuit classically. As such, characterization methods which fully describe the error profile, such as process tomography, are not practical for large systems. While Quantum Volume [44] has become very popular since its creation in 2019, it is not scalable for large quantum systems, and therefore is only suitable for the NISQ era.

As we enter into the *Quantum Discovery Regime*, where quantum computations are no longer classically simulable, we must ensure that these characterization techniques test circuits with outputs that can be calculated classically. In addition, state preparation and measurement (SPAM) errors can be calculated and corrected by other means. Therefore, it is important that the figure of merits calculated from these characterization protocols only indicate the effectiveness of the gates and as such are independent of SPAM.

One such family of protocols which meet these qualifications is Randomized Benchmarking. It is the most widely and rigorously studied characterization technique. In this thesis, we study this family of protocols to better understand their physical meaning under a variety of gate-sets, error models, and sample distributions. We then use this to create a standard which obtains the most relevant figures of merit while also minimizing the resources and time needed to obtain them.

The unitarity is a measure of the degree of coherence of noisy gates, and can be used to help identify whether the errors are caused by calibration errors (coherent) or by the system interacting with the environment (decoherent). It was developed by Wallman et al in [169] which created a novel protocol now renamed Extended Randomized Benchmarking (XRB). The authors established a rigorous theoretical framework for XRB under the physically unrealistic assumption that the noise is gate-independent. Chapter 3 extends the novel work done in [169] by establishing the physical meaning of unitarity under various gate-dependent error models by numerical simulations.

Standard randomized benchmarking (SRB) measures the rate of error, or infidelity, in the system. Recently, [131] showed that there is a gauge freedom between the average gate infidelity and the infidelity obtained by SRB. Chapter 4 addresses this disconnect by proving that for single-qubit experiments with high fidelity, the average gate infidelity and the infidelity obtained from SRB experiments differ due to a physical basis mismatch between the gates and the SPAM errors. This chapter then introduces a novel figure of merit which describes the expected average fidelity over random circuits composed of gates from a fixed gate-set known as the *gate-set circuit fidelity*, which is shown to be linearly related to the average infidelity of the gates in the gate-set, which is obtained via SRB.

Chapter 5 clarifies the standing conflation in the literature between the now standard Clifford-group RB (SRB) protocol proposed in [49, 45] and an alternate version of RB proposed later by NIST [92] (NIST RB), which non-uniformly samples from the Clifford group. In this chapter we discuss how NIST RB leads to an exponential fidelity decay, and that the infidelity measured between the two protocols compares by numerically simulating various gate-dependent error models.

Chapter 5 provides a perfect example of the importance of having a clear standard for characterizing basic gates-sets such that an impartial cross-platform comparison can be made. Chapter 6 lays out in detail the set of protocols that make up this standard, each of which provide different important figures of merit, that when used together, combine to give a well-rounded description of the error profile. We then make, for the first time, an even-handed cross-platform comparison of some of the world's leading quantum computing devices. This standard unbiasedly assesses basic gate-sets and their universality, for both standard and non-standard gates.

Chapter 2

Mathematical Background

2.1 Matrix preliminaries and inner products

Let A & B and their corresponding conjugate transposes $A^\dagger$ & $B^\dagger$ be any $d \times d$ matrices $(A, B, A^\dagger, B^\dagger \in M_d(\mathbb{C}))$. The matrix A can be normalized via

$$\underline{\mathrm{A}} \coloneqq \frac{A}{||A||}, \tag{2.1}$$

where the Frobenius norm [81] of A is

$$||A|| \coloneqq \sqrt{\langle A, A\rangle} = \sqrt{\mathrm{Tr}(A^\dagger A)}, \tag{2.2}$$

and $\langle A, A\rangle$ is the Hilbert-Schmidt inner product of A with itself.

Definition 1 (Hilbert-Schmidt inner product)**.** The overlap between the matrices A & B is given by the trace (Tr):

$$\langle A, B\rangle \coloneqq \mathrm{Tr}\big(A^\dagger B\big). \tag{2.3}$$

Definition 2 (Polar Decomposition)**.** There exists some unitary $d \times d$ matrix U such that for any matrix $A \in M_d(\mathbb{C})$

$$A = U|A|, \tag{2.4}$$

where the absolute value of matrix A is given by

$$|A| \coloneqq \sqrt{A^\dagger A} \tag{2.5}$$

Definition 3 (Commutator)**.** The commutator between the matrices A and B is

$$[A, B] \coloneqq AB - BA \tag{2.6}$$

where A and B are said to commute if $[A, B] = 0$ ($AB = BA$), and anti-commute if $AB = -BA$.

2.1.1 Vectorization of matrices

The vectorization of matrices will be used in section 2.2 in the representation of quantum states, as well as in section 2.7 to establish the superoperator formalism.

Any matrix $A \in M_d(\mathbb{C})$ can be expressed with respect to some orthonormal basis $\mathbb{B} = \{\underline{\mathrm{B}}_0, \dots, \underline{\mathrm{B}}_{d^2-1}\}$ [1] for $\underline{\mathrm{B}}_i = |i\rangle \langle i|$.

$$A = \sum_{i=0}^{d^2-1} \alpha_i \underline{\mathrm{B}}_i \tag{2.7}$$

where the constant (α_i) is the overlap between A and $\underline{\mathrm{B}}_i$:

$$\alpha_i = \langle A, \underline{\mathrm{B}}_i \rangle \tag{2.8}$$

Therefore we can use this to vectorize the matrix A:

$$|A\rangle\rangle_{\mathbb{B}} = \sum_{i=0}^{d^2-1} \alpha_i |i\rangle \tag{2.9}$$

for some orthonormal basis $|i\rangle$. The basis that will be used throughout this thesis when vectorizing quantum states will be the normalized Pauli basis.

Note that for two vectorized matricies $|w\rangle\rangle = \sum_i w_i |i\rangle$ and $|v\rangle\rangle = \sum_j v_j |j\rangle$ for the orthonormal basis $\{|i\rangle\}$ such that $\langle i| j\rangle = \delta_{ij}$:

$$\langle v| w\rangle = \langle v, w\rangle \tag{2.10}$$

---

[1] The basis elements $(\underline{\mathrm{B}}_i)$ should not be confused with the matrix B used throughout the section.

2.1.2 Special types of matrices

Normal, unitary, hermitian, and projection matrices will be frequently referred to throughout this thesis. We therefore provide a brief review of the definitions and properties of such matrices, and introduce the notation that will be used throughout.

Definition 4 (*Normal*)**.** A matrix (A) is Normal if

$$AA^\dagger = A^\dagger A. \tag{2.11}$$

This means that A and $A^\dagger$ commute, and that A is unitarily diagonalizable, ie there exists a unitary matrix (U) such that

$$D = UAU^\dagger$$

is a diagonal matrix.

Definition 5 (Unitary)**.** The matrix U is a $d \times d$ unitary matrix if:

$$UU^\dagger = U^\dagger U = \mathbb{I}. \tag{2.12}$$

Note that all unitary matrices are normal and $U^\dagger = U^{-1}$.

Definition 6 (*Hermitian or Self-Adjoint*)**.** A matrix A is Hermitian if

$$A^\dagger = A. \tag{2.13}$$

Note that if A is Hermitian it is also normal. Given that A has a discrete set of eigenvalues $\{\lambda_i\}$, it has a spectral decomposition:

$$A = \sum_i \lambda_i P_i, \tag{2.14}$$

where all eigenvalues $\{\lambda_i\}$ are real and $\{P_i\}$ are projectors onto the eigenspace of the corresponding distinct eigenvalues[69]. A rank-N projector (for an N-fold degenerate eigenvalue λ_i) can be expressed as:

$$P_i = \sum_{n_i=1}^{N} |i, n_i\rangle \langle n_i, i| \tag{2.15}$$

The expectation value of P_i gives the probability of the outcome λ_i

$$\Pr(\lambda_i) = \langle\psi| P_i |\psi\rangle \in [0, 1]. \tag{2.16}$$

Therefore the expectation value of A can be calculated using the Born rule probabilities [24]

$$\mathbb{E}(A) = \sum_i \lambda_i \Pr(\lambda_i). \tag{2.17}$$

Definition 7 (*Projectors*)**.** A set of matrices $\{P_i\}$ are projectors if

$$\begin{aligned} P_i &\geq 0, && \text{(Positive semi-definite)} \\ P_i P_j &= \delta_{ij} P_i, && \text{(Orthogonal)} \\ P_i &= P_i^\dagger, && \text{(Hermitian)} \\ P_i^2 &= P_i. && \text{(Idempotent)} \end{aligned}$$

Here $P_i \geq 0$ is the notation for a positive semi-definite operator, where $\langle\psi| P_i |\psi\rangle \geq 0 \; \forall \; |\psi\rangle \in \mathcal{H}$.

2.2 Quantum State

In this thesis quantum states will be used to describe the preparation procedure which can be represented by a vector in a finite Hilbert space. A qubit is a system whose observable are represented by the Pauli matrices. It is a quantum state which can be in some superposition of the two orthonormal states $|0\rangle$ and $|1\rangle$. A $d(= 2^n)$ dimensional quantum state can be represented by a density matrix ρ, which is a positive semi-definite self-adjoint matrix with unit trace:

$$\mathrm{Tr}(\rho) = 1 \tag{2.18}$$

A density matrix ρ can be a pure state or mixed state. Here an ideal preparation is denoted by a pure state

$$\rho = |\psi\rangle \langle\psi| , \tag{2.19}$$

while a mixed state is an ensemble of pure states,

$$\rho \equiv \sum_i p_i |\psi_i\rangle \langle\psi_i| , \tag{2.20}$$

where p_i is the probability of being in state $|\psi_i\rangle$.

Given a composite quantum system (ρ_{AB}) in the hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, the reduced operator of subsystem A $(\mathcal{H}_A)$ is

$$\rho_A \equiv \mathrm{Tr}_B \left(\rho_{AB}\right) . \tag{2.21}$$

Where the partial trace is given by:

$$\mathrm{Tr}_B(|a_1\rangle \langle a_2| \otimes |b_1\rangle \langle b_2|) \equiv |a_1\rangle \langle a_2| \,\mathrm{Tr}(|b_1\rangle \langle b_2|) \tag{2.22}$$

for any vectors $|a_1\rangle , |a_2\rangle$ & $|b_1\rangle , |b_2\rangle$ in the state spaces of A & B.

2.3 Measurement

A quantum measurement can be represented by a set of positive semi-definite measurement operators $\{\mu_m\}$ that sum to $\mathbb{I}$ for completeness. It is sometimes called a Positive Operator-Value Measure (POVM), where m is a observable event corresponding to the POVM element μ_m. An ideal measurement is a projector valued measure (PVM), which is a POVM in the special case where $\{\mu_m\}$ is a set of orthogonal projectors.

Using Born's rule, the probability of observing the event m given that the quantum state was in $\rho = |\psi\rangle\langle\psi|$ immediately before the measurement is:

$$\Pr(m|\rho) = \langle\psi|\,\mu_m\,|\psi\rangle = \langle\mu_m, \rho\rangle \tag{2.23}$$

For completeness, the probability of all observable events must be equal to 1,

$$1 \equiv \sum_m \Pr(m|\rho) = \sum_m \langle\psi|\,\mu_m\,|\psi\rangle\,, \tag{2.24}$$

therefore $\sum_m \mu_m = \mathbb{I}$.

2.4 Quantum Operation

A quantum operation (or quantum channel[2]), is a linear map which takes quantum states to quantum states via a physical transformation. Specifically, it is a completely positive trace preserving (CPTP) linear map.

For any state ρ_{AB} in the composite system $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, a map Φ acting on $\mathcal{H}_A$ ($\Phi : L(\mathcal{H}_A) \rightarrow L(\mathcal{H}_A)$) is defined as a quantum channel iff Φ is a completely positive, trace-preserving, and linear. A definition for each of these properties follows.

Definition 8 (*Completely Positive*)**.** Φ is completely positive if $\Phi \otimes \mathbb{I}_B$ is positive

$$\Phi \otimes \mathbb{I}_B(\rho_{AB}) \geq 0. \tag{2.25}$$

Definition 9 (*Trace Preserving*)**.** Φ is trace-preserving if

$$\operatorname{Tr}\Phi(\rho_A) = \operatorname{Tr}\rho_A = 1 \tag{2.26}$$

---

[2] Note that "quantum operation", "quantum channel", and "quantum gate" will be used interchangeably in this thesis.

Definition 10 (*Convex Linearity*)**.** For any given state that has a probability p_i of the state being in ρ_i, a channel Φ is said to be a linear map if:

$$\Phi\left(\sum_i p_i \rho_i\right) = \sum_i p_i \Phi(\rho_i) \tag{2.27}$$

Ideal quantum operations are unitary channels, which represent transformations in closed quantum systems. In practice these gates are applied in an open system, as the external system brings energy in and out of the system. The focus of this thesis is noise in (the more physically realistic) open quantum systems which cause quantum operations to be implemented non-ideally. [3]

Definition 11 (*Unitary*)**.** A channel ($\mathcal{U}(\rho) = U\rho U^\dagger$) is said to be unitary iff:

$$\mathcal{U}^\dagger \circ \mathcal{U}(\rho) = \mathcal{U} \circ \mathcal{U}^\dagger(\rho) = \rho \ . \tag{2.28}$$

Specifically,

$$U^\dagger U = UU^\dagger = \mathbb{I} \tag{2.29}$$

where $U^\dagger = U^{-1}$, and thus unitary operations are reversible.

2.4.1 Important Quantum Channels

The gates introduced in this section are common in quantum algorithms, and will reappear throughout this thesis.

Single Qubit Gates

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

An arbitrary pulse along any Cartesian axis system will be denoted as:

$$P(H, \theta) := e^{i\theta H/2} \ . \tag{2.30}$$

Note, in chapter 5, for simplicity, the Pauli Pulses will be denoted as:

$$X_\theta := e^{-i\theta/2\ X},\ Y_\theta := e^{-i\theta/2\ Y},\ Z_\theta := e^{-i\theta/2\ Z},$$

---

[3] Note that unitary noise in closed systems does exist, and will be discussed further in the coming chapters.

Entangling Gates

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \; CZ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \; SWAP = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2.5 Twirling and Unitary t-designs

Definition 12. A *unitary t-design* is a finite set of N unitary matrices ($\{U_n\}$) such that for any given polynomial ($P_{t,t}(U)$) of degree of at most t in both matrix elements of U, and their complex conjugates:

$$\frac{1}{N}\sum_{n=1}^{N} P_{t,t}(U_n) = \int_{\mathcal{U}(d)} P_{t,t}(U)d\mu(U) \tag{2.31}$$

Here, averaging $P_{t,t}(U)$ over all of $\mathcal{U}(d)$ is equivalent to averaging over the set $\{U_n\}$.

Definition 13. The *twirl* of a channel (Φ) over a discrete unitary 2-design ($\boldsymbol{U}_i \in \boldsymbol{U}$),

$$\begin{aligned} \boldsymbol{\Phi}^{\boldsymbol{U}}(\rho) &= \sum_{i=1}^{|\boldsymbol{U}|} p_i \boldsymbol{U}_i^{\dagger} \circ \boldsymbol{\Phi} \circ \boldsymbol{U}_i(\rho) \\ &= \sum_{i=1}^{|\boldsymbol{U}|} p_i \left(U_i^{\dagger} \boldsymbol{\Phi}(U_i \rho U_i^{\dagger}) U_i \right) \end{aligned} \tag{2.32}$$

The twirl of Φ over the Haar measure is a depolarizing channel.

2.6 Gate Sets

This section will cover the gate-sets that are used throughout this thesis.

2.6.1 Group

A group $\mathcal{G}$ is a set of elements $\{\mathcal{G}_i\}$ that satisfy:

1. Closure: $\mathcal{G}_i \cdot \mathcal{G}_j = \mathcal{G}_k, \mathcal{G}_k \in \mathcal{G}$
2. Identity Element: $\exists \mathbb{I} \in \mathcal{G}\ s.t.\ \mathbb{I} \cdot \mathcal{G} = \mathcal{G}, \forall \mathcal{G} \in \mathcal{G}$
3. Associativity: $\mathcal{G}_i \cdot (\mathcal{G}_j \cdot \mathcal{G}_k) = (\mathcal{G}_i \cdot \mathcal{G}_j) \cdot \mathcal{G}_k$
4. Inverse Element: $\forall \mathcal{G} \in \mathcal{G}, \exists \mathcal{G}^{-1} \in \mathcal{G},\ s.t.\ \mathcal{G} \cdot \mathcal{G}^{-1} = \mathcal{G}^{-1}\mathcal{G} = \mathbb{I}$

2.6.2 Pauli

The n-qubit Pauli group ($\mathbb{P}_n$) is a 1-design and is defined as:

$$\mathbb{P}_n = \{\mathbb{I}, X, Y, Z\}^{\otimes n} \times \{\pm 1, \pm i\} \tag{2.33}$$

2.6.3 Cliffords

The n-qubit Clifford group $\mathbb{C}_n$ is a 2-design that is a normalizer of the Pauli group ($\mathbb{P}_n$):

$$C \in \mathbb{C}_n \iff CPC^\dagger \in \mathbb{P}_n,\ \forall P \in \mathbb{P}_n \tag{2.34}$$

Any single-qubit unitary with this property is an element of $\mathbb{C}$ up to some global phase $e^{i\phi}$. The 24 single qubit Clifford gates ($\mathbb{C}_1$) can be decomposed into a series of $\pi/2$ and π pulses about X, Y, Z. Similarly, the 11,520 2-qubit Clifford gates ($\mathbb{C}_2$) can be constructed using $\mathbb{C}_1$ and an entangling gate[4]. Please see the supplementary material of [10] for their full decomposition descriptions.

2.6.4 SU(4)

Special Unitary group (SU(d)) is a set of all $d \times d$ unitary matrices with determinant=1, where $d = 2^n$. For example, for d=4,

$$SU(4) = \{A = 4 \times 4 \text{ complex matrix} \mid A^\dagger A = \mathbb{I},\ \det(A) = 1\} \tag{2.35}$$

Due to the global phase freedom, all unitary quantum operations are equivalent to an element of SU(d).

Note that $\mathbb{P}_n \subset \mathbb{C}_n \subset SU(d) \subset \mathbb{U}_n$.

---

[4] This entangling gate must generate the CNOT gate with $\mathbb{C}_1$ (CNOT, CZ, SWAP, iSWAP...)

2.7 Superoperator Formalism

Quantum errors are ubiquitously present in quantum systems, and in many cases cause the system of interest to interact with the environment (open quantum system). This includes the loss of information, or loss or gain of energy from the system to the environment. This section will discuss how these noisy operators which act on open quantum systems can be represented using the superoperator formalism.

2.7.1 Kraus Operator Sum Decomposition

Any CPTP map can be represented using the Kraus operator sum decomposition

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger, \tag{2.36}$$

where the linear operators A_i are the kraus operators. Since $\mathbf{\Phi}$ is trace preserving,

$$\sum_i A_i^\dagger A_i = \mathbb{I}. \tag{2.37}$$

The Kraus operator sum decomposition for a unitary channel is

$$\mathbf{\Phi}(\rho) = U \rho U^\dagger.$$

for $U \in \mathbb{U}$.

2.7.2 Chi Matrix Representation

For a fixed basis set of operators $\mathbb{B}$ on the same state space, the Kraus operators can be expressed in terms of basis elements as

$$A_i = \sum_m \alpha_{im} \underline{\mathrm{B}}_m \tag{2.38}$$

where the constant (α_{im}) is defined as the overlap between the Kraus operator A_i and the basis element $\underline{\mathrm{B}}_m$:

$$\alpha_{im} = \mathrm{Tr}\big(\underline{\mathrm{B}}_m^\dagger A_i\big) \tag{2.39}$$

Therefore equation 2.36 becomes:

$$\Phi(\rho) = \sum_{mn} \chi_{mn} \underline{\mathrm{B}}_m \rho \underline{\mathrm{B}}_n^{\dagger} \tag{2.40}$$

Where χ has matrix elements:

$$\chi_{mn} \equiv \sum_{i} \alpha_{im} \alpha_{in}^{*} \tag{2.41}$$

Note that:

$$\chi_{00} = \frac{1}{d^2} \sum_{i} |\operatorname{Tr}(A_i)|^2 \tag{2.42}$$

For trace-preserving channels $\chi_{00} = 1$, which implies that it is a measure of the leakage/loss.

2.7.3 Liouville (Process Matrix)

The Liouville representation, which will be widely used throughout this thesis, is useful when working with sequences of channels because the composition of channels is given by matrix multiplication. Since a channel acts on a state by multiplying a vectorization of the state by the Liouville representation of the channel, we can write, for example,

$$\Phi_2 \circ \Phi_1(\rho) \cong \boldsymbol{\Phi}_2 \boldsymbol{\Phi}_1 |\rho\rangle\rangle \tag{2.43}$$

To obtain the vectorization of a state used by the Liouville representation, we start with an orthonormal basis,

$$\mathbb{B} = \{\underline{\mathrm{B}}_1, \ldots, \underline{\mathrm{B}}_{d^2}\} \forall \underline{\mathrm{B}}_i \in \mathbb{C}^{d^2}, \tag{2.44}$$

where,

$$\langle \underline{\mathrm{B}}_i, \underline{\mathrm{B}}_j \rangle = \delta_{i,j} \ \forall \ \underline{\mathrm{B}}_i, \underline{\mathrm{B}}_j \in \mathbb{B}.$$

A density matrix ρ can then be expressed as:

$$\rho = \sum_{i \in \mathbb{N}_{d^2}} \langle \underline{\mathrm{B}}_i, \rho \rangle \underline{\mathrm{B}}_i.$$

Therefore, the Liouville representation of ρ is a column vector $|\rho\rangle\rangle \in \mathbb{C}^{d^2}$ where the i^{th} element is $\langle \underline{\mathrm{B}}_i, \rho\rangle$.

For example, let's express the density matrix $\rho = |0\rangle\langle 0|$ in the Pauli Liouville representation[5] $(|\rho\rangle\rangle)$.

$$\rho = |0\rangle\langle 0| = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = 1/2\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + 1/2\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} = 1/2\mathbb{I} + 1/2Z$$

therefore,

$$|\rho\rangle\rangle = 1/2\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\begin{array}{c}(\mathbb{I})\\(X)\\(Y)\\(Z)\end{array}$$

Note that the ordering ([$\mathbb{I}$,X,Y,Z]) for states and matrices is used throughout this thesis. The Liouville representation of a channel Φ with respect to the basis $\mathbb{B}$ is a unique matrix $\mathbf{\Phi} \in \mathbb{C}^{d^2\times d^2}$ such that,

$$\mathbf{\Phi}|\rho\rangle\rangle = |\Phi(\rho)\rangle\rangle \tag{2.45}$$

with matrix elements;

$$[\mathbf{\Phi}]_{ij} = \langle \underline{\mathrm{B}}_i, \Phi(\underline{\mathrm{B}}_j)\rangle = \langle\langle \underline{\mathrm{B}}_i|\mathbf{\Phi}|\underline{\mathrm{B}}_j\rangle\rangle.$$

Therefore the composition of channels can be represented by matrix multiplication. Similarly, let's express the Hadamard (H) in the Pauli Liouville representation $\boldsymbol{H}$.

$$H\mathbb{I}H^\dagger = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \mathbb{I}$$

Similarly;

$$HXH^\dagger = Z \qquad HYH^\dagger = -Y \qquad HZH^\dagger = X$$

Therefore we can write $\boldsymbol{H}$ as:

$$\begin{array}{c} (\mathbb{I})(X)(Y)(Z) \\ \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & -1 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}\begin{array}{c}(\mathbb{I})\\(X)\\(Y)\\(Z).\end{array}\end{array}$$

---

[5] Pauli Liouville representation is the Liouville representation given that the Pauli group is the orthonormal basis that is used.

If Φ is a CP map than it can be represented in block form as:

$$\mathbf{\Phi} = \begin{pmatrix} S(\Phi) & \vec{\mathbf{\Phi}}_{sdl} \\ \vec{\mathbf{\Phi}}_n & \mathbf{\Phi}_u \end{pmatrix}. \tag{2.46}$$

Here, $S(\Phi)$ is the average survival rate (scalar), $\vec{\mathbf{\Phi}}_{sdl}$ is the state-dependent leakage (1×4^n-1), and $\vec{\mathbf{\Phi}}_n$ and $\mathbf{\Phi}_u$ are the non-unital (4^n-1 $\times$ 1) and unital parts (4^n-1 $\times$ 4^n-1). Note that if $\mathbf{\Phi}$ is a trace preserving map, then $S(\Phi) = 1$ and $\vec{\mathbf{\Phi}}_{sdl} = \vec{0}$. Also, if $\mathbf{\Phi}$ is unital than $S(\Phi) = 1$, and $\vec{\mathbf{\Phi}}_n = \vec{\mathbf{\Phi}}_{sdl} = \vec{0}$.

2.8 Distance Measures

2.8.1 Distance measures between quantum states

Definition 14. The *Trace Distance* between the quantum states ρ and σ is:

$$D(\rho, \sigma) \equiv \frac{1}{2} \operatorname{tr} |\rho - \sigma| \tag{2.47}$$

For qubits, physically, this value is half of the Euclidean distance between the vectors of ρ and σ on the bloch sphere.

Definition 15. The *fidelity* of the states ρ and σ is:

$$f(\rho, \sigma) \equiv \operatorname{tr} \sqrt{\rho^{1/2} \sigma \rho^{1/2}} \tag{2.48}$$

The fidelity is a symmetric comparison of two states, with a value that ranges between 0 and 1. The fidelity is 1 iff the states are identical, and 0 if the states have support on mutually orthogonal subspaces.

Note that the Trace Distance between ρ and σ can be bounded by the fidelity as [61]:

$$1 - \sqrt{f(\rho, \sigma)} \leq D(\rho, \sigma) \leq \sqrt{1 - f(\rho, \sigma)} \tag{2.49}$$

The upper bound is reached in the case were both ρ and σ are both pure states, and in the case where ρ and/or σ is a pure state, then the lower bound is tightened to:

$$1 - f(\rho, \sigma) \leq D(\rho, \sigma) \tag{2.50}$$

2.8.2 Fidelity between quantum operations

Definition 16. The *average gate fidelity* of the noisy experimental implementation ($\tilde{\mathcal{G}}$) of an ideal unitary channel ($\mathcal{G}$) is:

$$F(\tilde{\mathcal{G}}, \mathcal{G}) \equiv \int d\psi \, \mathrm{Tr}\Big[\mathcal{G}^\dagger(\psi)\tilde{\mathcal{G}}(\psi)\Big] \tag{2.51}$$

This figure of merit is an average of the closeness of the operations over all physical pure states, and thus is independent of the input state.

The noisy channel can be expressed as $\tilde{\mathcal{G}} = \boldsymbol{\mathcal{E}} \circ \mathcal{G}$ where $\boldsymbol{\mathcal{E}}$ is the error caused by implementing the gate $\mathcal{G}$ (in the ideal case $\boldsymbol{\mathcal{E}} = \mathbb{I}$). Therefore, eq. (2.51) can be rewritten as [175]:

$$F(\boldsymbol{\mathcal{E}}) \equiv F(\boldsymbol{\mathcal{E}}, \mathbb{I}) = \frac{1}{d+1} + \frac{\mathrm{tr}(\boldsymbol{\mathcal{E}})}{(d^2+d)} \tag{2.52}$$

Note that the fidelity cannot distinguish between coherent and decoherent errors, but instead is a figure of merit of the overlap between the expected operation, and the noisy one that is actually applied.

Definition 17. The *infidelity* of an error channel ($\boldsymbol{\mathcal{E}}$) is an indicator of how far $\boldsymbol{\mathcal{E}}$ is from $\mathbb{I}$ instead of how close. It is defined in terms of F as:

$$r(\boldsymbol{\mathcal{E}}) = 1 - F(\boldsymbol{\mathcal{E}}) \tag{2.53}$$

Definition 18. The *process fidelity* is a figure of merit defined by equation eq. (2.42) and can be written in terms of F as:

$$\chi_{00}(\boldsymbol{\mathcal{E}}) = \frac{(d+1)F(\boldsymbol{\mathcal{E}}) - 1}{d} \tag{2.54}$$

Definition 19. The *process infidelity* of an error channel ($\boldsymbol{\mathcal{E}}$) written in terms of χ_{00} as:

$$e_F(\boldsymbol{\mathcal{E}}) = 1 - \chi_{00}(\boldsymbol{\mathcal{E}}) \tag{2.55}$$

Definition 20. The *effective depolarizing parameter* (p) of an error channel $\boldsymbol{\mathcal{E}}$ in the form of eq. (2.46) is an indicator of leakage/loss which amounts to a shrinkage of the Bloch sphere:

$$p(\boldsymbol{\mathcal{E}}) = \frac{\mathrm{Tr}\, \boldsymbol{\mathcal{E}}_u}{d^2 - 1} \tag{2.56}$$

	F	r	p	χ_{00}	e_F
F	F	$1-r$	$\frac{(d-1)p+1}{d}$	$\frac{d\chi_{00}+1}{d+1}$	$1-\frac{d}{d+1}e_F$
r	$1-F$	r	$\frac{d-1}{d}(1-p)$	$\frac{d}{d+1}(1-\chi_{00})$	$\frac{d}{d+1}e_F$
p	$\frac{dF-1}{d-1}$	$1-\frac{d}{d-1}r$	p	$\frac{d^2\chi_{00}-1}{d^2-1}$	$1-\frac{d^2}{d^2-1}e_F$
χ_{00}	$\frac{(d+1)F-1}{d}$	$1-\frac{d+1}{d}r$	$\frac{(d^2-1)p+1}{d^2}$	χ_{00}	$1-e_F$
e_F	$\frac{d+1}{d}(1-F)$	$\frac{d+1}{d}r$	$\frac{d^2-1}{d^2}(1-p)$	$1-\chi_{00}$	e_F

Table 2.1: This table from [34] is a summary of the linear relationships between the fidelity (F from eq. (2.52)), infidelity (r from eq. (2.53)), effective depolarizing parameter (p from eq. (2.56)), process fidelity (χ_{00} from eq. (2.42)), and process infidelity (e_F from eq. (2.55))

These figures of merit can be adjusted such that we redefine $\boldsymbol{\mathcal{E}}$ to be the average error for a set of operations $\mathbb{G} = \{\mathcal{G}_i\}$ such that:

$$\boldsymbol{\mathcal{E}}^{\mathbb{G}} = \mathbb{E}_{\mathcal{G}\in\mathbb{G}}(\mathcal{G}^{\dagger}\tilde{\mathcal{G}}) \tag{2.57}$$

Therefore, the average gate-set fidelity of the gate-set $\mathbb{G}$ is defined as $F(\boldsymbol{\mathcal{E}}^{\mathbb{G}})$. As we will see in section 2.10, these figures of merit can be estimated efficiently using variants of the Randomized Benchmarking protocol.

2.8.3 Unitarity

The unitarity [169] of a channel is its degree of coherence. Coherent errors tend to positively interfere with each other in a sequence and thus cause worse case errors than their decoherent counterparts. This means that coherent errors have a larger impact on the fidelity. Therefore, the unitarity is an indicator of how much of the error is due to miscalibration, and how much is from a loss of energy or information about the system to the environment.

Definition 21. *Unitarity* (u) of the error channel ($\boldsymbol{\mathcal{E}}$) is defined as the average purity of the output states averaged over all pure states ($\phi = |\psi\rangle\langle\psi|$):

$$u(\boldsymbol{\mathcal{E}}) = \frac{d}{d-1}\int d\phi\,\mathrm{Tr}\,\boldsymbol{\mathcal{E}}(\phi - \mathbb{I}/d)^2$$

The unitarity can also be written in terms of the unital part of the noise by:

$$u(\boldsymbol{\mathcal{E}}) = \frac{1}{d^2-1} \operatorname{Tr} \boldsymbol{\mathcal{E}}_u^\dagger \boldsymbol{\mathcal{E}}_u \tag{2.58}$$

Where $u(\boldsymbol{\mathcal{E}}) = 1$ for any purely unitary error and $u(\mathcal{D}) = p^2$ for a purely depolarizing channel $\mathcal{D}$.

2.8.4 Bound on the worst case error

Definition 22. The *diamond norm* of a channel Φ is:

$$||\Phi||_\diamond \equiv \sup_\psi ||\mathbb{I} \otimes \Phi(\psi)|| \tag{2.59}$$

The worst case error rate ($\frac{1}{2}||\mathcal{E} - \mathbb{I}||_\diamond$) is an important figure of merit as it takes into account the effects of error propagating to and from the ancilla's in the environment. It gives an upper bound on the impact an error can have on a quantum system, in comparison to the fidelity which gives the average/expected impact. Unfortunately it cannot be efficiently measured due to its maximization over all states. Therefore, it is more practical to find the bounds on the diamond norm, which are a function of the infidelity (r) as seen in [175]:

$$\frac{(d+1)r(\mathcal{E})}{d} \leq \frac{1}{2}||\mathcal{E} - \mathbb{I}||_\diamond \leq \sqrt{d(d+1)r(\mathcal{E})} \tag{2.60}$$

These bounds were improved in [170], by accounting for how much of the error is coherent:

$$\frac{C(\mathcal{E})}{\sqrt{2}} \leq ||\mathcal{E}||_\diamond \leq \sqrt{\frac{d^3 C(\mathcal{E})^2}{4} + \frac{(d+1)^2 r(\mathcal{E})^2}{2}} \tag{2.61}$$

for,

$$C(\mathcal{E})^2 = \frac{d^2-1}{d^2}(u(\mathcal{E}) - 2p(\mathcal{E}) + 1). \tag{2.62}$$

From this bound it is clear that when the error is coherent, both the upper and lower bounds scale as $\mathcal{O}(\sqrt{r(\mathcal{E})})$ (since $u(\mathcal{E}) = 1$), and when the error is purely stochastic, ie:

$$u(\mathcal{E}) \approx \left(1 - \frac{dr(\mathcal{E})}{d-1}\right)^2,$$

then the bounds scale as $\mathcal{O}(r(\mathcal{E}))$. Note that this is the case for any norm-based measure, and is not a feature of the "worst-case" nature of the diamond norm. In [170], Wallman showed this by looking at the average instead of the maximum over all input states, and saw this behaviour.

2.9 Types of Noise

This thesis focuses on the characterization of noise in the form of errors that are ubiquitously present in open quantum systems. Note that we look at non-adversarial noise, and specifically we assume that the error is Markovian. This section gives a brief description of those errors, as well as their super-operator representations.

2.9.1 Coherent Errors

The upper bound on the infidelity (or worst case error) of a composition of gates is caused by the buildup of coherent errors, as coherent errors positively (and sometimes negatively) interfere in sequence. As we will see in chapter 5, this behaviour causes the infidelity to increases non-linearly with respect to the sequence length.

Coherent errors are unitary channels ($\mathcal{U}(\rho)$) where the inverse of the channel ($\mathcal{U}^{-1}(\rho)$) is its dual ($\mathcal{U}^{\dagger}(\rho)$)

$$\mathcal{U} \circ \mathcal{U}^{\dagger}(\rho) = \mathcal{U}^{\dagger} \circ \mathcal{U}(\rho) = \rho, \tag{2.63}$$

and can be written in the form

$$\mathcal{U}(\rho) = U \rho U^{\dagger}, \tag{2.64}$$

where U is the single unitary Kraus operator. Unitary error ($\mathcal{E}$) happens when a target operator U is miscalibrated to a slightly different unitary operator U', where $\mathcal{E} = U'U^{\dagger}$. This produces a rotation on the surface of the bloch sphere, without causing shrinkage.

A common gate-dependent unitary error (that will be referred to extensively in this thesis) is the over-rotation of a target operator. In general, for an ideal gate $\mathcal{G} = P(H, \theta)$ (from eq. (2.30)), over-rotation can be represented by $\tilde{\mathcal{G}} = P(H, \theta + \alpha)$.

For example, let the target be the Hadamard gate:

$$H = \frac{1}{\sqrt{2}} (X + Z) \tag{2.65}$$

therefore using Taylor Expansion, and the fact that $H^2 = \mathbb{I}$:

$$\begin{aligned} e^{iH\theta} &= \left(iH\theta - \frac{i}{3!}(H\theta)^3 + \frac{i}{5!}(H\theta)^5 - ...\right) + \left(\mathbb{I} - \frac{(H\theta)^2}{2!} + \frac{(H\theta)^4}{4!} - ...\right) \\ &= iH\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - ...\right) + \mathbb{I}\left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - ..\right) \\ &= iH\sin\theta + \mathbb{I}\cos\theta \end{aligned}$$

Noting that:

$$H = e^{iH(\pi/2)}$$

then for a given over-rotation parameter α:

$$\begin{aligned} e^{iH(\pi/2+\alpha)} &= iH(\sin(\pi/2+\alpha) + \mathbb{I}\cos(\pi/2+\alpha)) \\ &= iH\cos\alpha - \mathbb{I}\sin\alpha \end{aligned}$$

removing a $-i$ global phase

$$= i\mathbb{I}\sin\alpha + \frac{(X+Z)}{\sqrt{2}}\cos\alpha$$

Using the approach from section 2.7.3, the Pauli-Liouville representation of an over-rotated Hadamard gate is given by:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sin^2\alpha & -\sqrt{2}\cos\alpha\sin\alpha & \cos^2\alpha \\ 0 & \sqrt{2}\cos\alpha\sin\alpha & (1-2\cos^2\alpha) & -\sqrt{2}\cos\alpha\sin\alpha \\ 0 & \cos^2\alpha & \sqrt{2}\cos\alpha\sin\alpha & \sin^2\alpha \end{bmatrix}$$

2.9.2 Cross-talk

An important type of coherent error process, called *cross-talk* [179, 142, 147, 2, 155, 161], is defined as a process which does not satisfy locality and/or independence of local operations

[148]. These conditions are defined as follows: The *locality condition* is met when a circuit does not couple qubits or subsets of qubits unless it contains operations that are intended to do so and the *independence condition* is met when each operation acting on each subset of qubits is unaffected by operations that are simultaneously applied to other subsets of qubits.

One of the most common types of crosstalk is spatial addressing errors [128, 122, 164, 28, 105]. These errors can occur between qubits, or between subsets of qubits. For example, an error could be introduced if an X rotation is applied to qubit A, and the pulse used to implement that rotation causes a small unintended X rotation on qubit B. In general, by driving qubit A, the drive field also affects qubit B and therefore violates the independence condition. An operation will satisfy the locality condition if a semiclassical description is sufficient to describe the operation as there will then be no direct coupling between the qubits.

Another type of crosstalk, referred to as “idle crosstalk” in [148], violates locality and can be seen most clearly when all qubits aren’t subject to any driving field. For example, if there is an always-on entangling gate error, each qubit’s state would depend on the state of the other qubits and would thus violate the locality condition as it is not a tensor product map.

We might want coupling to always exist for the implementation of a CR gate (controlled phase shift). However, when driving single-qubit gates, this coupling can induce nonlocal crosstalk. Similarly, if qubit A is coupled to B and B to C, then when implementing a CR gate on A and B, the BC coupling may create non-local crosstalk. Both of these scenarios are examples of a kind of crosstalk called static coupling which violates the locality condition.

2.9.3 Incoherent Errors

An incoherent error is a process by which information from the quantum system is lost to the environment, which causes the bloch sphere to shrink. Unlike coherent errors, incoherent errors cannot be easily corrected by a simple calibration, but in sequence they do not lead to worst case errors.

Amplitude Damping

Amplitude damping (or T1 decay) is the relaxation of a quantum system which causes energy (and consequently phase information) to be lost from the system to the environment.

This error causes the Bloch sphere to contract towards the north pole (ground state).

$$(r_x, r_y, r_z) \rightarrow (r_x\sqrt{1-\gamma}, r_y\sqrt{1-\gamma}, \gamma + r_z(1-\gamma)) \tag{2.66}$$

One representation of the amplitude damping Kraus operators are

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \quad A_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix},$$

where the decay constant ($\gamma \in [0,1]$) represents the probability that the system spontaneously emits a photon. The Pauli-Liouville representation of a single-qubit amplitude damping channel is

$$\boldsymbol{\mathcal{E}}_\gamma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & \sqrt{1-\gamma} & 0 \\ \gamma & 0 & 0 & 1-\gamma \end{bmatrix},$$

which has a fidelity of

$$F(\boldsymbol{\mathcal{E}}_\gamma) = 2/3 + (2\sqrt{1-\gamma} - \gamma)/6.$$

Phase Damping

Phase Damping (or T2*/T2 decay) is the loss of quantum information without the loss of energy. The phase contains the information about the state (X and Y), while the amplitude (Z) corresponds to the amount of energy. Thus the bloch sphere is squeezed about the X and Y axis, while the Z axis is unaffected ($(r_x, r_y, r_z) \rightarrow (r_x(1-\lambda), r_y(1-\lambda), r_z)$). The Kraus operators for this channel are

$$A_0 = \sqrt{1-\lambda/2}\ \mathbb{I}, \quad A_1 = \sqrt{\lambda/2}\ Z,$$

for a decay constant $\lambda \in [0,1]$. Phase damping has a Pauli-Liouville representation of

$$\boldsymbol{\mathcal{E}}_\lambda = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\lambda & 0 & 0 \\ 0 & 0 & 1-\lambda & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

with a fidelity of

$$F(\boldsymbol{\mathcal{E}}_\lambda) = 1 - \lambda/3.$$

Depolarizing

Depolarizing noise is the physical process where with a probability $(1-p)$ all information about the system is lost (state becomes the maximally mixed state $\mathbb{I}$), and with probability p, the state remains intact.

$$\mathcal{D}(\rho) = \frac{(1-p)\mathbb{I}}{d} + p\rho \tag{2.67}$$

Here we define $p \in [0,1]$ as the effective depolarizing constant. This type of error causes the Bloch sphere to uniformly contract (as a function of p) towards the origin $((r_x, r_y, r_z) \to (1-p)(r_x, r_y, r_z))$. The Kraus operators associated with this error are given by

$$A_0 = \sqrt{(1+3p)/4}\mathbb{I}, \quad A_1 = \sqrt{(1-p)/4}X, \quad A_2 = \sqrt{(1-p)/4}Y, \quad A_3 = \sqrt{(1-p)/4}Z.$$

With a Pauli-Liouville representation of

$$\boldsymbol{\mathcal{E}}_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{bmatrix},$$

and fidelity

$$F(\boldsymbol{\mathcal{E}}_p) = \frac{1+p}{2}. \tag{2.68}$$

2.10 Error Characterization

Quantum Characterization, Verification, & Validation (QCVV) is the field of study for estimating the quality of quantum informational processors [168]. QCVV can be broken into 4 distinct categories which include characterizing quantum states, measurements, processes, and holistic benchmarks. This thesis will focus on Randomized Benchmarking like protocols which make up a sub-field of quantum process characterization.

2.10.1 Quantum State

The first type of assessment confirms the production of a quantum state with desirable properties. This includes Quantum State Tomography (QST) which was first proposed in 1968 [63]. Since 1989, QST has been used extensively to both demonstrate and quantify how accurately the state of interest is produced [140, 165, 176, 82, 120, 84, 1, 38]. This type of assessment motivates variants of QST such as QST via compressed sensing[66] from 2010 which reduced the overhead of this assessment, QST of the permutationally invariant part of the density operator[162] from 2010, a self-learning technique from 2014 [56], and QST for many-body systems [97] from 2017.

2.10.2 Quantum Measurement

The second type of assessment is the characterization of quantum measurements and is the least developed area of study out of the 4 categories presented here. The most widely used method for characterizing measurement error is similar to QST, where the user prepares and then measures various quantum states to find the overlap between the expected and measured outcomes [104, 15]. As such, this type of characterization strongly depends on the state preparation error. Measurement and state characterization are related by a gauge transformation, so it is very difficult to differentiate between state preparation and measurement errors. For this reason state preparation and measurement errors are usually talked about together as SPAM.

2.10.3 Quantum Process

The third category of QCVV evaluates quantum operations, and is the focus of this thesis. This type of assessment of quantum devices can be split into three categories; 1. non-scalable protocols which obtain partial information about the noisy effective identity gate and assumes insignificant SPAM errors, 2. non-scalable protocols which fully characterize the noisy process, and 3. protocols which partially characterize the noise and are robust to SPAM.

Non-Scalable, SPAM dependent protocols which obtain partial information about the noisy effective identity gate

The first generation of QCVV experiments were developed for nuclear magnetic resonance (NMR). The first method proposed was to measure physical properties of devices and compare the expected values of measurement outcomes between devices. Fundamentally, what these protocols attempt to do is characterize the identity operator. The first implementation of this type of assessment was the Rabi oscillation experiment in 1937 [136]. This experiment is used to calculate the relaxation time (T_1) of the qubit from the excited state to the ground state. In 1950, the Ramsey oscillation experiment [137] was expanded to calculate the lifetime of coherence under both reversible (dephasing contribution from relaxation time) and irreversible dephasing noise (T_2^*). Also in 1950, the spin-echo[68] experiment was proposed to estimate the lifetime of coherence under only irreversible dephasing noise (T_2).

Non-scalable protocols which obtain complete information about noisy channels

Assume insignificant SPAM error: The first protocol which assessed quantum operations called Quantum Process Tomography (QPT) was created in 1997[129, 40], and has since been developed further and used extensively[58, 124, 177, 102, 39, 163, 88, 87]. QPT aims to reconstruct the noisy process by preparing a set of quantum states, applying the process to them, and then using QST to identify the output states. The downside to fully reconstructing the process is that the overhead is very large, it takes a long time to implement, and the outcome is dependent on SPAM. In 2003, an ancilla-assisted version of QPT was developed [5], and in 2004, Hamiltonian tomography [150] was created to experimentally identify the Hamiltonian of a controlled 2-level system. In 2005, [80, 125] showed that as an alternative to QPT, in order to minimize overhead, complimentary classical fidelities could be used to partially characterize the gates.

In 2006, [117] obtained a direct characterization of the quantum dynamics which relies on approaches for error-detection which did not implement QST. In 2009, [18] constructed a way to obtain the Lindbladian generators using symplectic tomography. [144] (2012) proposed an efficient estimation method of the Pauli channel parameters via Pauli channel tomography. In a continued effort to reduce the overhead of QPT-like protocols, in 2014, [138] created compressed sensing process tomography.

SPAM robust: In 2013, Gate Set Tomography (GST) was developed [115, 20, 90, 141, 178, 126, 151, 36, 139, 109, 134, 100, 143, 51] to reconstruct a set of processes, similar to implementing QPT on each gate in a gate-set, in a self consistent way. Standard GST (i.e. GST without model reduction) is not scalable, although its cost may be drastically reduced by incorporating model-based approaches [123, 21].

SPAM independent protocols which obtain partial information about the error model

Scalable in the high fidelity regime: The protocols presented here are scalable only in the high fidelity regime, as the twirling gates are each composed of many basic gates. The high fidelity regime is defined by the number of samples (circuits & shots) to get a fixed relative precision on p, where the number of samples needed[6] grows as p^{-10} for low p. This causes the exponential to become harder to resolve as the system size increases, if not within the high fidelity regime.

In 2005, Standard Randomized Benchmarking from protocol 1 (SRB) was developed with the use of SU(d)[49], which obtained the average gate fidelity of the SU(d) gate-set. This evaluation is independent of SPAM, scalable, and fast to implement when compared to QPT-like protocols. Unlike QPT, SRB only obtains partial information about the system instead of reconstructing the full process. Recent works have utilized different gate-sets than SU(d) to probe other aspects of the error model[45, 92, 10, 32, 43, 77, 60, 27, 133, 74]. In 2011 and 2012, [106, 107] supported the robustness of the SRB protocol under essentially arbitrary, non-adversarial state preparation and measurement errors. Interleaved Randomized Benchmarking (IRB) protocol was developed by [108] (2012), which enabled the error estimate of a single gate of interest. In 2015, [90] developed a protocol to estimate the phase of the universal single-qubit gate-set in a robust and efficient way. [153] developed Iterative Randomized Benchmarking (2016) which is similar to Interleaved Randomized Benchmarking but instead implements the gate of interest many times between the twirling gates, in order to amplify its error contribution relative to the twirling gate error. In 2017 and 2018 the robustness of SRB-like protocols was supported by rigorous proofs that the observed error rate relates directly to a well-defined notion of gate-fidelity

---

[6] For low p, the number of samples needed is $N = 100/(A^2 p^{10})$. For example, for N on the order of millions (10^6), and A of order 1, than $p \approx 0.4$.

[171, 131, 31], which fully overcame concerns about relating the measured RB error rates to a meaningful concept of gate-fidelity under gate-dependent errors [131]. The systematic error bounds on the estimates from IRB were tightened in 2019 [34] by accounting for the amount of coherent error in the system.

Scalable: In 2019, Cycle Benchmarking from protocol 4 (CB) [52] was developed which enabled the characterization of n-qubit cycles of large circuits efficiently. K-body Noise Reconstruction [59, 71] is a variant of CB, developed in 2020, which is used to estimate the probability distribution of the errors affecting the system, and can detect all correlated Pauli errors. This work was further developed in 2021, when [72] created a protocol similar but distinct from [59] with a more efficient recovery guarantee.

Non-Scalable: In 2015, [169] designed an SRB-like protocol, called Extended Randomized Benchmarking from protocol 3, which estimated the stochastic contribution to the average-gate fidelity, enabling the user to differentiate between calibration and stochastic noise. This protocol is deemed non-scalable because with larger system sizes, the number of circuits becomes too large to obtain fixed precision.

2.10.4 Holistic Benchmarking

Holistic Benchmarking aims to test the quantum device as a whole, by calculating volumetric figures of merit, testing the device's ability to perform specific tasks, or in some cases how well the device performs that task relative to its classical counterpart.

This includes testing the devices ability to experimentally implement algorithms such as Shor's algorithm [154] starting in 2007 [96, 118], Grover search algorithm [67] starting in 2009[48, 57], Deutsch-Jozsa algorithm [47] in 2009 [48, 101], and the variational quantum eigensolver (VQE) algorithm [127] beginning in 2018 [78].

Randomized Compiling [174] converts coherent and non-Markovian error into stochatic error by randomizing the gates in a circuit using virtual twirling groups, which significantly reduces run-time error. When CB [52] is implemented in conjunction with Randomized Compiling, the fidelity of any sized circuit can be estimated, and thus any algorithm.

Volumetric figures of merit include Quantum Volume (2017) [19, 44] which calculates the largest square circuit (number of qubits = circuit length) that the device can implement with reasonable accuracy, and volumetric benchmarking (2020) [22, 132] which is similar to quantum volume, but defines a family of rectangular circuits (number of qubits $\neq$ circuit length). Cross-entropy benchmarking (2019) [6, 8], which belongs to the family of RB-like protocols, also aims to address this type of assessment, and specifically is used to estimate the largest random circuit that can be performed sufficiently well to exceed the performance of a classical noisy sampling. An explicit connection between Cross-Entropy benchmarking (2019) and the general framework of RB is made in [76].

Applications which are hard to classically simulate and can be used for supremacy comparisons include quantum chemistry applications [149] (2016), Quantum Approximate Optimization Algorithm [53] (2016), boson sampling [98, 103] (2016), IQP (instantaneous quantum polynomial-time) circuits [25] (2017), translation-invariant Ising spin model [65] (2017), certifying ground states of frustration-free Hamiltonians based on simple energy measurements of local Hamiltonian terms [70] (2017), and cross-entropy benchmarking [6, 8] (2019).

2.10.5 Randomized Benchmarking and QCVV

QCVV is tending towards protocols which assess quantum processes and holistic benchmarks. Specifically, in the future, when we approach the end of the NISQ era we will need to assess how well a device is able to implement computations, with particular focus on those applications that classical computing devices are unable to implement efficiently.

A member of the Randomized Benchmarking family can be represented by a protocol which leads to the sum of exponential fidelity decays, is SPAM independent, and uses randomization. As can be seen by the discussion above, protocols in the Randomized Benchmarking family can fall in two categories, namely 1) scalable (or scalable in high fidelity regime), SPAM independent partial process characterization, and 2) holistic benchmarks.

The Randomized Benchmarking family has continued to grow to accommodate for the challenges presented by the error models (cross-talk, gate-dependent,...), device requirements, compilation tools (RC), and the ways to assess quantum computing devices. It is a mature family of protocols, which are supported by rigorous proofs and theoretical underpinnings, which will continue to adapt and grow to accommodate the future challenges of the assessment of quantum computing hardware.

Protocols that belong to the Randomized Benchmarking family are used to obtain figures of merit which can be used to infer the reliability of a quantum device. These protocols only obtain partial information about the system, and are not only scalable but also fast to run in comparison to Tomography approaches. RB-like protocols can be used to obtain information about the amount of coherent error in the device vs the amount of stochastic error, and can be used for Noise Reconstruction[59]. In addition, RB protocols are also independent of SPAM error, so the figures of merit obtained by them give estimates of just the gate operation error. Some disadvantages of this type of characterization are that these types of protocols do not fully characterize the noise (as is seen in gate-set tomography), do not characterize SPAM, and can only be used to characterize Markovian non-catastrophic[30] errors.

The main complementary approaches to the family of RB protocols include Tomography-like protocols, Ramsey and Rabi experiments, and Quantum Volume. The advantage of the Tomography like protocols over RB-like protocols are that they are able to fully characterize the error of the process or gate-set and they can estimate the SPAM. The disadvantages are that they are slow to implement, have large overheads, and they cannot always separate the SPAM from the gate error. The Ramsey and Rabi experiments give clear information about T_1 and T_2 times, although they do not provide any additional insight into the coherent error, SPAM, cross-talk, or many other aspects of the error. Unlike Cross-Entropy Benchmarking, Quantum Volume does not attempt to illustrate supremacy, but instead is used to quantify the maximum size of a circuit the device could implement before surpassing a mathematically defined error threshold. This is done by sampling circuits for which the outcome distribution can be computed classically, and then slowly increasing the circuit depth and width until the observed distribution is considered too noisy according to a technical criterion. Therefore this type of Benchmark is not scalable as the outcomes need to be classically simulable.

Quantum Volume, Cross-entropy benchmarking, Gate-Set Tomography, and Volumetric benchmarking are examples of protocols which can be used for making cross-platform comparisons. In chapter 6, we present an alternative to these approaches which focuses on the error of the basic gate-sets. We show that this technique allows us to obtain an even-handed cross-platform comparison, and the use of both Clifford or non-Clifford basic gate-sets tests the device's ability to implement a universal gate-set (which are the primitives for Quantum Volume).

The benefits to this type of assessment (chapter 6) in comparison to the alternatives (presented above) is that it is scalable, fast to implement, and obtains information about various aspects of the error model which capture the devices essential performance features. The major weakness to this type of assessment is that it does not obtain complete information about the error profile, and thus does not pinpoint the physical mechanism responsible for the error. The main challenges that remain for this type of cross-platform comparison is to extend it to also test the scalability of the device, i.e. implement Cycle Benchmarking on various qubit pairs of the device.

2.10.6 Randomized Benchmarking Protocols

The four protocols we present here are based on the same family of benchmarking protocols called Randomized Benchmarking. The first is called Standard Randomized Benchmarking (SRB) and it calculates the average gate fidelity of the given gate-set. Note that this section is just to introduce the reader to the protocols that will be considered throughout this thesis, and will be discussed more thoroughly in the following chapters.

Note also that this protocol description has been transcribed from [23]/ chapter 5, and has been placed here for convenience.

Protocol 1: SRB, as described in [45, 49].

---

1. Sample a set of m gates G_i picked independently and uniformly at random from the 2-design $\mathbb{G}$
2. Determine the recovery gate G_{m+1} (can be set to $G_{m:1}^{-1}$);
3. Prepare a state $\rho = |0\rangle\langle 0|$;
4. Perform the sampled gates from step 1, followed by the recovery gate G_{m+1} determined in step 2:
   $\tilde{G}_{m+1:1} = \tilde{G}_{m+1} \circ \ldots \circ \tilde{G}_1$;
5. Measure a POVM $\{Q, \mathbb{I} - Q\}$, where the first observable is $Q \approx G_{m+1:1}(|0\rangle\langle 0|)$, and respective outcome labels are {"recovery", "non-recovery"};
6. Repeat steps 3–5 a number times to estimate the probability of observing the "recovery" event $\Pr(\text{“recovery”}\,|\{G_i\}, m) = \operatorname{tr} Q\tilde{G}_{m+1:1}(\rho)$;
7. Repeat steps 1–6 for s different sets of m randomly sampled gates $\{G_i\}$;
8. Repeat steps for 1–7 for different values of m of random gates.
9. Fit the estimated recovery probabilities to the decay model
   $$A_{\mathbb{G}}\, p_{\mathbb{G}}^m + B_{\mathbb{G}}\ ; \tag{2.69}$$
10. Estimate the Clifford gate-set infidelity through
    $$r_{\mathbb{G}} = \frac{d-1}{d}(1 - p_{\mathbb{G}}) \tag{2.70}$$
    for $d = 2^n$ for an n-qubit system.

The second protocol is Interleaved Randomized Benchmarking (IRB). It is very similar to SRB except it calculates the average infidelity (over $\mathbb{C}$) of the chosen gate of interest composed with a random Clifford. It can be used along with SRB to approximate the single gate error of the gate of interest.

Protocol 2: IRB, as described in [108].

---

1. Sample a set of m gates G_i picked independently and uniformly at random from the 2-design $\mathbb{G}$

2. Determine the recovery gate G_{m+1} which is the inverse of the composition of the m random gates, including the interleaved C gate;

3. Prepare a state $\rho \approx |0\rangle\langle 0|$;

4. Perform the sampled gates from step 1, each followed by the interleaved gate C, followed by the recovery gate G_{m+1} determined in step 2: $\tilde{G}_{m+1:1} = \tilde{G}_{m+1} \circ \tilde{C} \circ \tilde{G}_m \circ \ldots \circ \tilde{C} \circ \tilde{G}_1$;

5–8. Idem as in protocol 1.

9. Fit the estimated recovery probabilities to the decay model

$$A_{\overline{C}}\, p_{\overline{C}}^m + B_{\overline{C}}\ ; \tag{2.71}$$

10. Estimate the gate error using the value $p_{\mathbb{G}}$ obtained by running

$$r_C^{est} = \frac{d-1}{d}\left(1 - \frac{p_{\overline{C}}}{p_{\mathbb{G}}}\right)\ . \tag{2.72}$$

The third protocol calculates the amount of the error which is incoherent, and can be used along with SRB to calculate the coherent infidelity.

Protocol 3: XRB, as described in [175] .

1. Choose a random sequence of m gates from the (Clifford) group $\mathbb{C}$:

$$\mathbf{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_m)$$

2. Prepare the state ρ (usually $|0\rangle\langle 0|$)

3. Apply the sequence:

$$\tilde{\mathbf{G}} = \tilde{\mathcal{G}}_m \tilde{\mathcal{G}}_{m-1} \ldots \tilde{\mathcal{G}}_1$$

4. Estimate the expectation value of Q given the sequence $\mathbf{G}$ is applied:

$$Q_{\mathbf{G}} = \langle\langle \boldsymbol{Q} | \, \tilde{\boldsymbol{\mathcal{G}}}_m \ldots \tilde{\boldsymbol{\mathcal{G}}}_1 \, | \boldsymbol{\rho} \rangle\rangle \tag{2.73}$$

5. Repeat steps 1-4 to find the expected average over all sequences of length m:

$$\mathbb{E}_{\mathbf{G}}\left[Q_{\mathbf{G}}^2\right] = \frac{1}{|\mathbb{C}|^m} \sum_{\boldsymbol{G} \in \mathbb{C}^{\boldsymbol{m}}} Q_{\mathbf{G}}^2 \tag{2.74}$$

where $|\mathbb{C}|^m$ is all of the possible sequences of m gates from the Clifford group $\mathbb{C}$. This expression can be rewritten as:

$$\mathbb{E}\left[Q_{\mathbf{G}}^2\right] = \langle\langle \boldsymbol{Q_2} | \left(\frac{1}{|\mathbb{C}|} \sum_{\boldsymbol{\mathcal{G}} \boldsymbol{\epsilon} \mathbb{C}} \tilde{\boldsymbol{\mathcal{G}}}^{\otimes 2} \right)^{\mathbf{m}} | \boldsymbol{\rho}^{\otimes \mathbf{2}} \rangle\rangle \tag{2.75}$$

where we measure along the X, Y and Z axes:

$$|\boldsymbol{Q_2}\rangle\rangle = |\boldsymbol{X}^{\otimes 2}\rangle\rangle + |\boldsymbol{Y}^{\otimes 2}\rangle\rangle + |\boldsymbol{Z}^{\otimes 2}\rangle\rangle.$$

Note that we are now taking the average over the gates in the group U, instead of over all the possible sequences.

6. Repeat steps 1-5 for different number of gates (m) and fit to:

$$\mathbb{E}\left[Q_{\mathbf{G}}^2\right] = A + B u_{fit} \left(\frac{1}{|\mathbb{C}|} \sum_{\boldsymbol{\mathcal{G}} \boldsymbol{\epsilon} \mathbb{C}} \tilde{\boldsymbol{\mathcal{G}}} \right)^{m-1} \tag{2.76}$$

to estimate the value of u_{fit}.

By use of equation 2.58, we can calculate the unitarity of gate dependent error and compare it to this fitted value

The final protocol is called Cycle Benchmarking (CB). Unlike the other protocols, CB uses the Pauli group as the twirling group. For a system with more than 1 qubit, Pauli error is significantly lower than Clifford error because the Clifford gates are composed of entangling gates, which generally lead to larger error than single qubit Pauli's. Similar to IRB, CB is used to calculate the inferred infidelity of the clock cycle (or interleaved gate in the case of IRB), although due to the very low error of the Pauli's it is significantly better at accurately predicting it. In addition to this, when used with Randomized Compiling, it can be used to predict the error rate of an entire protocol, not just a single cycle or gate.

Protocol 4: CB, as described in [52].

---

1. Select a set of N-qubit Pauli gates of length $|\mathbb{P}|$.
2. Choose two sequence lengths m_1 and m_2, s.t. $\mathcal{G}^{m_2-m_1}$ is factorizable ($\mathcal{G}_1' \otimes \mathcal{G}_2'$).
3. For each Pauli matrix $P \in \mathbb{P}$, L random sequences per Pauli for $l \in (1, \ldots, L)$, and sequence lengths $m \in (m_1, m_2)$, implement the sequence;

$$\tilde{\mathcal{C}}(P) = \tilde{\mathcal{R}}_m \tilde{\mathcal{G}} \tilde{\mathcal{R}}_{m-1} \tilde{\mathcal{G}} \ldots \tilde{\mathcal{R}}_1 \tilde{\mathcal{G}} \tilde{\mathcal{R}}_0 \tag{2.77}$$

   where $\mathcal{R}_j$ are N-qubit Pauli cycles. Then after calculating the expected outcome $\mathcal{C}(P)$, estimate their overlap via:

$$f_{P,m,l} = \mathrm{tr}\left[\mathcal{C}(P)\tilde{\mathcal{C}}(\rho)\right] \tag{2.78}$$

   for an initial state ρ that is a $+1$ eigenstate of P.

4. The composite process fidelity can then be estimated via:

$$F_{RC}(\tilde{\mathcal{G}}, \mathcal{G}) = \sum_{P \in \mathbb{P}} \frac{1}{|\mathbb{P}|} \left(\frac{\sum_{l=1}^{L} f_{P,m_2,l}}{\sum_{l=1}^{L} f_{P,m_1,l}} \right)^{\frac{1}{m_2-m_1}} . \tag{2.79}$$

Note that the ratio of the dressed $\mathcal{G}$ (what CB measures) and the local fidelities of the Pauli gates is used to estimate the process fidelity of $\mathcal{G}$, ie:

$$\frac{F_{RC}(\tilde{\mathcal{G}}, \mathcal{G})}{F_{RC}(\tilde{\mathcal{I}}, \mathcal{I})} \tag{2.80}$$

Chapter 3

Unitarity under Gate-Dependent noise

3.1 Overview

We extend the work done by [169] that provided a method for estimating the degree of coherence in the noise acting on a system via a variation on randomized benchmarking protocols, which was done with the underlying assumption that noise is independent of the target gate. While this might be a good approximation when the errors vary slightly from operation to operation, it will likely not be accurate when the errors vary substantially over the set of operations. In this work, we explore what this protocol measures when this assumption is relaxed. In particular we look at what it measures when errors vary over the gate set, to determine if it tracks the average unitarity over the gate set, or the unitarity of the average. This is important to be able to distinguish because the average unitarity of each error is able to distinguish between gate dependent coherent and decoherent errors, while the unitarity of the average error is not. This was done by analytically proving that the output should be the average unitary of each gate-dependent error. By examining gate-dependent error models for the single-qubit Clifford gates, we numerically show that under physically realistic gate-dependent errors, the figure of merit obtained via XRB overlaps strongly with the average unitarity of the each of the gate-dependent errors, and has little overlap with the unitarity of the average over the gate-dependent error.

Introduction

One of the largest challenges to the development of a quantum computer is its sensitivity to the errors ubiquitously present in controlled quantum systems, arising, for example, due to their coupling to the quantum environment. Thankfully it has been proven that physically reasonable error models can be corrected, enabling robust quantum computation, provided that the error does not surpass a code-dependent threshold [3, 91, 93], known as the fault-tolerant threshold. However, we must ensure that our diagnostic methods require fewer classical computational resources than simply simulating an ideal quantum circuit, otherwise the quantum computer provides no practical computational advantage. One method of efficiently estimating noise is to consider benchmarking circuits using elements from a fixed gate set under the assumption that the errors are gate-independent[169].

Recent work has clarified how certain kinds of quantum errors, such as coherent noise, will generally lead to a larger worst-case error than stochastic noise (see section 2.8.4) [107, 175, 146], and can be corrected in different ways[55]. Therefore the ability to efficiently and robustly measure whether the noise is coherent or not will greatly affect our ability to characterize a given experimental approach to fault-tolerant quantum computation. In Ref. [169], an efficient protocol for estimating the coherence of noise was introduced and proven to be robust to state preparation and measurement errors under the assumption that the errors are gate-independent. However, it is problematic to assume that noise takes the form of gate-independent coherent errors because one of the most natural sources of coherent error depends on the gate.

In 2015, [169] proposed a method to characterize the degree of coherence of the error. This method was later renamed extended randomized benchmarking (XRB). This was an extraordinary finding as the returned figure of merit, called the unitarity, quantifies what portion of the error was due to coherent calibration errors, and what portion was due to the system interacting with the environment. Calibration errors are generally easier to correct and lead to larger worst case errors than their decoherent counterparts. Therefore the unitarity gives crucial information to the user about how much error will remain in the system in the limit where they are able to correct all of the calibration errors. This can help experimentalists decide whether that level of error is acceptable or if they need to take the next, much more challenging step, of redesigning their system to reduce these decoherent errors.

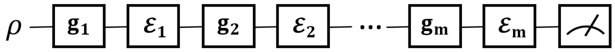
One of the main drawbacks to this revolutionary characterization protocol is that the rigorous theoretical framework built to support the protocol, was developed under the physically unrealistic assumption that the noise is gate independent. This chapter will discuss the effectiveness of XRB when the noise is gate-dependent, and will give a physical interpretation of the figure of merits that it produces. The two types of gate-dependent error that will be used to study the unitarity in this chapter are gate-dependent over-rotations of Clifford gates and amplitude damping.

Gate Dependent XRB

In comparison to the XRB protocol presented in protocol 3, consider the case when the error is gate dependent. In this scenario, the sequence of gates ($\boldsymbol{\mathcal{G}}_i$) and their corresponding error ($\boldsymbol{\mathcal{E}}_i$) becomes

$$\boldsymbol{\mathcal{E}}_m\boldsymbol{\mathcal{G}}_m\boldsymbol{\mathcal{E}}_{m-1}\boldsymbol{\mathcal{G}}_{m-1}\dots\boldsymbol{\mathcal{E}}_1\boldsymbol{\mathcal{G}}_1,$$

and can be represented in circuit form as,



Relabeling $\tilde{\boldsymbol{\mathcal{G}}}_j = \boldsymbol{\mathcal{E}}_j\boldsymbol{\mathcal{G}}_j$, under any gate-dependent Markovian noise model, equation 2.75 becomes

$$\mathbb{E}\left[Q_{\mathbb{G}}^2\right] = \langle\langle \boldsymbol{Q_2}| \left(|\mathbb{G}|^{-1} \sum_{\boldsymbol{\mathcal{G}}\epsilon\mathbb{G}} \tilde{\boldsymbol{\mathcal{G}}}^{\otimes 2} \right)^m |\boldsymbol{\rho}^{\otimes 2}\rangle\rangle \tag{3.1}$$

$$= \langle\langle \boldsymbol{Q_2}|\boldsymbol{\mathcal{M}}^m|\boldsymbol{\rho}^{\otimes 2}\rangle\rangle \tag{3.2}$$

for $\boldsymbol{\mathcal{M}} = |\mathbb{G}|^{-1}\sum_{\boldsymbol{\mathcal{G}}\epsilon\mathbb{G}}\tilde{\boldsymbol{\mathcal{G}}}^{\otimes 2}$, and $|\boldsymbol{Q_2}\rangle\rangle = |\boldsymbol{X}^{\otimes 2}\rangle\rangle + |\boldsymbol{Y}^{\otimes 2}\rangle\rangle + |\boldsymbol{Z}^{\otimes 2}\rangle\rangle$.

We will use first order perturbation theory to calculate the eigenvalues ($\tilde{\lambda}$) of $\boldsymbol{\mathcal{M}}$. First we decompose $\boldsymbol{\mathcal{M}}$ via:

$$\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}}_0 + \delta\boldsymbol{\mathcal{M}} \tag{3.3}$$

where $\delta\boldsymbol{\mathcal{M}}$ is the perturbation matrix and $\boldsymbol{\mathcal{M}}_0$ is the ideal implementation of $\boldsymbol{\mathcal{M}}$,

$$\boldsymbol{\mathcal{M}}_0 = |\mathbb{G}|^{-1}\sum_{\boldsymbol{\mathcal{G}}\epsilon\mathbb{G}}\boldsymbol{\mathcal{G}}^{\otimes 2}.$$

$\boldsymbol{\mathcal{M}}_0$ has a degenerate space with eigenvalues of $\lambda_{0_{\{1,2\}}} = 1$, and eigenvectors $v_{0_1} = |\mathbb{I}_{d^2}\rangle\rangle$ and $v_{0_2} = |\mathcal{S}\rangle\rangle$, where $\mathcal{S}$ is the SWAP gate.

Then we can use 1st order perturbation theory, where

$$\begin{aligned}\lambda_{1_i} &= v_{0_i}^{\dagger}\boldsymbol{\mathcal{M}}v_{0_i} &(3.4)\\ &= \lambda_{0_i} + \delta\lambda_{1_i}. &(3.5)\end{aligned}$$

for the small perturbation $\delta\lambda_{1_i}$, which corresponds to a small error acting on $\tilde{\mathcal{G}}$.

Let $|B_1\rangle\rangle$ and $|B_2\rangle\rangle$ be the orthonormal basis vectors which span the 2 trivial irreducible representations (irreps).

$$\begin{aligned}B_1 &= \mathbb{I}_{d^2}/d &(3.6)\\ B_2 &= (\mathcal{S} - \mathbb{I}_{d^2}/d)/\sqrt{d^2-1} &(3.7)\end{aligned}$$

Therefore, calculating the 1st order eigenvalues of $\boldsymbol{\mathcal{M}}$ via eq. (3.4) and the orthonormal basis vectors from eqs. (3.6) and (3.7) [169],

$$\begin{aligned}\langle\langle\mathbb{I}_{d^2}|\boldsymbol{\mathcal{M}}|\mathbb{I}_{d^2}\rangle\rangle &= \mathrm{Tr}\left[|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathbb{I}_{d^2}\tilde{\mathcal{G}}^{\otimes 2}(\mathbb{I}_{d^2})\right]\\ &= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathrm{Tr}\left[\tilde{\mathcal{G}}^{\otimes 2}(\mathbb{I}_{d^2})\right]\\ &= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathrm{Tr}\left[\tilde{\mathcal{G}}(\mathbb{I}_{d})\right]^2\\ &= d^2|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}S(\tilde{\mathcal{G}})^2 &(3.8)\end{aligned}$$

where here $S(\tilde{\boldsymbol{\mathcal{G}}})$ is the survival error rate from eq. (2.46).

$$
\begin{aligned}
\langle\langle \mathbb{I}_{d^2}|\boldsymbol{\mathcal{M}}|\mathcal{S}\rangle\rangle = \langle\langle \mathcal{S}|\boldsymbol{\mathcal{M}}^\dagger|\mathbb{I}_{d^2}\rangle\rangle &= \mathrm{Tr}\left[|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathcal{S}\tilde{\mathcal{G}}^{\dagger\otimes 2}(\mathbb{I}_{d^2})\right] \\
&= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathrm{Tr}\left[\mathcal{S}\tilde{\mathcal{G}}^{\dagger\otimes 2}(\mathbb{I}_{d^2})\right] \\
&= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathrm{Tr}\left[\tilde{\mathcal{G}}^{\dagger}(\mathbb{I}_d)\tilde{\mathcal{G}}^{\dagger}(\mathbb{I}_d)\right] \\
&= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}d\left|\left|\tilde{\mathcal{G}}_{sdl}\right|\right|^2 + dS(\tilde{\mathcal{G}})^2 \qquad (3.9)
\end{aligned}
$$

where $\tilde{\mathcal{G}}_{sdl}$ is the state-dependent leakage from eq. (2.46) of $\tilde{\mathcal{G}}$. Similarly,

$$
\begin{aligned}
\langle\langle \mathcal{S}|\boldsymbol{\mathcal{M}}|\mathbb{I}_{d^2}\rangle\rangle &= \mathrm{Tr}\left[|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathcal{S}\tilde{\mathcal{G}}^{\otimes 2}(\mathbb{I}_{d^2})\right] \\
&= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathrm{Tr}\left[\mathcal{S}\tilde{\mathcal{G}}^{\otimes 2}(\mathbb{I}_{d^2})\right] \\
&= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathrm{Tr}\left[\tilde{\mathcal{G}}(\mathbb{I}_d)\tilde{\mathcal{G}}(\mathbb{I}_d)\right] \\
&= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}d\left|\left|\tilde{\mathcal{G}}_{n}\right|\right|^2 + dS(\tilde{\mathcal{G}})^2 \qquad (3.10)
\end{aligned}
$$

where $\tilde{\mathcal{G}}_n$ is the non-unital part of $\tilde{\mathcal{G}}$ from eq. (2.46) and is very small.

Lastly, given that $\mathcal{S} = \sum_k A_k \otimes A_k^\dagger$, where A_k forms an orthonormal operator basis[1],

---

[1] One such set of Kraus operators are $A_k \in \mathbb{P}_1$

$$\begin{aligned}\langle\langle\mathcal{S}|\boldsymbol{\mathcal{M}}|\mathcal{S}\rangle\rangle &= \mathrm{Tr}\left[|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\mathcal{S}\tilde{\mathcal{G}}^{\otimes 2}(\mathcal{S})\right]\\ &= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\sum_{k}\mathrm{Tr}\left[\mathcal{S}\tilde{\mathcal{G}}(A_k)\otimes\tilde{\mathcal{G}}(A_k^\dagger)\right]\\ &= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\sum_{k}\mathrm{Tr}\left[\tilde{\mathcal{G}}(A_k)^\dagger\tilde{\mathcal{G}}(A_k)\right]\\ &= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\left|\left|\tilde{\mathcal{G}}\right|\right|^2 \end{aligned} \tag{3.11}$$

Therefore using the expressions above,

$$\begin{aligned}\boldsymbol{\mathcal{M}}_{11} &= \langle\langle\frac{\mathbb{I}_{d^2}}{d}|\boldsymbol{\mathcal{M}}|\frac{\mathbb{I}_{d^2}}{d}\rangle\rangle\\ &= |\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}S(\tilde{\mathcal{G}})^2,\end{aligned} \tag{3.12}$$

$$\begin{aligned}\boldsymbol{\mathcal{M}}_{12} &= \langle\langle\frac{\mathbb{I}_{d^2}}{d}|\boldsymbol{\mathcal{M}}|\frac{\mathcal{S}}{\sqrt{d^2-1}}\rangle\rangle - \langle\langle\frac{\mathbb{I}_{d^2}}{d}|\boldsymbol{\mathcal{M}}|\frac{\mathbb{I}_{d^2}}{d\sqrt{d^2-1}}\rangle\rangle\\ &= (d^2-1)^{-1/2}|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\left|\left|\tilde{\mathcal{G}}_n\right|\right|^2\end{aligned} \tag{3.13}$$

$$\begin{aligned}\boldsymbol{\mathcal{M}}_{21} &= \langle\langle\frac{\mathcal{S}}{\sqrt{d^2-1}}|\boldsymbol{\mathcal{M}}|\frac{\mathbb{I}_{d^2}}{d}\rangle\rangle - \langle\langle\frac{\mathbb{I}_{d^2}}{d\sqrt{d^2-1}}|\boldsymbol{\mathcal{M}}|\frac{\mathbb{I}_{d^2}}{d}\rangle\rangle\\ &= (d^2-1)^{-1/2}|\mathbb{G}|^{-1}\sum_{\mathcal{G}\epsilon\mathbb{G}}\left|\left|\tilde{\mathcal{G}}_n\right|\right|^2\end{aligned} \tag{3.14}$$

Using the definition from eqs. (2.2) and (2.46), we see that $\left|\left|\tilde{\mathcal{G}}\right|\right|^2 = S(\tilde{\mathcal{G}})^2 + \left|\left|\tilde{\mathcal{G}}_{sdl}\right|\right|^2 + \left|\left|\tilde{\mathcal{G}}_n\right|\right|^2 + \left|\left|\tilde{\mathcal{G}}_u\right|\right|^2$, where $\tilde{\mathcal{G}}_u$ is the unital part of $\tilde{\mathcal{G}}$, and thus;

$$
\begin{aligned}
\boldsymbol{\mathcal{M}}_{22} &= \langle\langle \frac{\mathcal{S}}{\sqrt{d^2-1}} | \boldsymbol{\mathcal{M}} | \frac{\mathcal{S}}{\sqrt{d^2-1}} \rangle\rangle + \langle\langle \frac{\mathbb{I}_{d^2}}{d\sqrt{d^2-1}} | \boldsymbol{\mathcal{M}} | \frac{\mathbb{I}_{d^2}}{d\sqrt{d^2-1}} \rangle\rangle \\
&\quad - \langle\langle \frac{\mathcal{S}}{\sqrt{d^2-1}} | \boldsymbol{\mathcal{M}} | \frac{\mathbb{I}_{d^2}}{d\sqrt{d^2-1}} \rangle\rangle - \langle\langle \frac{\mathbb{I}_{d^2}}{d\sqrt{d^2-1}} | \boldsymbol{\mathcal{M}} | \frac{\mathcal{S}}{\sqrt{d^2-1}} \rangle\rangle \\
&= (d^2-1)^{-1} |\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} \left|\left| \tilde{\mathcal{G}} \right|\right|^2 - S(\tilde{\mathcal{G}})^2 - \left|\left| \tilde{\mathcal{G}}_n \right|\right|^2 - \left|\left| \tilde{\mathcal{G}}_{sdl} \right|\right|^2 \\
&= (d^2-1)^{-1} |\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} \left|\left| \tilde{\mathcal{G}}_u \right|\right|^2 \qquad (3.15) \\
&= |\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} u(\tilde{\mathcal{G}}) \qquad (3.16)
\end{aligned}
$$

Where $u(\tilde{\mathcal{G}})$ is the unitarity of the noisy gate $\tilde{\mathcal{G}}$.

Therefore, from equation eq. (3.4), the eigenvalues of $\boldsymbol{\mathcal{M}}$ are:

$$
\lambda_{1_1} = \boldsymbol{\mathcal{M}}_{11} = |\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} S(\tilde{\mathcal{G}})^2 \qquad (3.17)
$$

which is equal to 1 if the gate-dependent noise is trace preserving (making $\delta\lambda_{1_1} \approx 0$), and

$$
\begin{aligned}
\lambda_{1_2} = \boldsymbol{\mathcal{M}}_{22} &= (d^2-1)^{-1} |\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} \left|\left| \tilde{\mathcal{G}}_u \right|\right|^2 \qquad (3.18) \\
&= |\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} u(\tilde{\mathcal{G}}) \qquad (3.19)
\end{aligned}
$$

In the high fidelity regime ($|\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} u(\tilde{\mathcal{G}}) \approx 1$), and therefore $\delta\lambda_{1_2} \approx 0$. Following the logic from [169] that $\boldsymbol{\mathcal{M}}$ is diagonalizable:

$$
\begin{aligned}
\mathbb{E}_j(Q_j^2) &= \langle\langle Q^2 | \boldsymbol{\mathcal{M}}^m | \rho^{\otimes 2} \rangle\rangle \\
&= A\lambda_{1_1}^m + B\lambda_{1_2}^m \qquad (3.20) \\
&= A + B|\mathbb{G}|^{-1} \sum_{\mathcal{G} \epsilon \mathbb{G}} u(\tilde{\mathcal{G}}) \qquad (3.21)
\end{aligned}
$$

for trace-preserving noise (ie $S(\tilde{\mathcal{G}}) = 1$).

Therefore, we expect that the output from the XRB protocol (u_{fit}) to be the average unitary of each gate-dependent error:

$$u_{ave} = \frac{1}{|\mathbb{G}|} \sum_{\boldsymbol{\mathcal{G}} \epsilon \mathbb{G}} u\left(\tilde{\boldsymbol{\mathcal{G}}} \circ \boldsymbol{\mathcal{G}}^{-1}\right).$$

We will also numerically show that in the gate dependent case, the average unitary of each error is not equivalent to the unitarity of the average error of the gate set, defined as:

$$u(\boldsymbol{\mathcal{E}}_{ave}) = u\left(\frac{1}{|\mathbb{G}|} \sum_{\boldsymbol{\mathcal{G}} \epsilon \mathbb{G}} \tilde{\boldsymbol{\mathcal{G}}} \circ \boldsymbol{\mathcal{G}}^{-1}\right).$$

for $u(x)$ defined in eq. (2.58).

These values are calculated via equation 2.58, and then compared to the numerically simulated fit of the XRB experiment via equation 2.76.

3.2 Numerical Results

In this section, we explore the relationship between the unitarity of the average error ($u(\mathcal{E}_{ave})$) and the average of the unitarities of each error (u_{ave}) via a numerical comparison in which two gate-dependent error scenarios are considered, over-rotation and amplitude dampening. For each gate a small random $\alpha, \gamma \in [0, 0.1]$ are assigned as in sections 2.9.1 and 2.9.3. In fig. 3.1 the two errors are applied simultaneously to each gate, while in fig. 3.2 the errors are considered separately.

The unitarity u_{fit} obtained via protocol 3 is compared against the unitarity of the average error ($u(\boldsymbol{\mathcal{E}}_{ave})$) and the average of the unitarities of each error (u_{ave}) where:

$$\Delta u(\mathcal{E}_{ave}) = |u_{fit} - u(\mathcal{E}_{ave})| \tag{3.22}$$

and

$$\Delta u_{ave} = |u_{fit} - u_{ave}|. \tag{3.23}$$

We plot the results of this comparison in fig. 3.1, where we see that once there are enough gates applied to create a good fit (≈ 10), the average of the unitarity of each gate

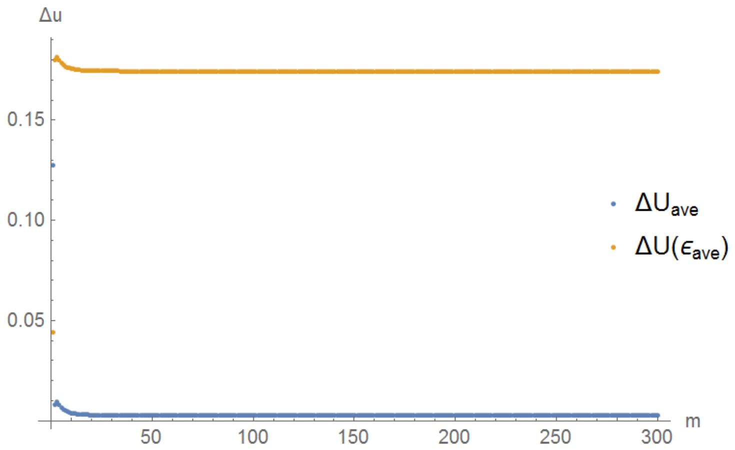


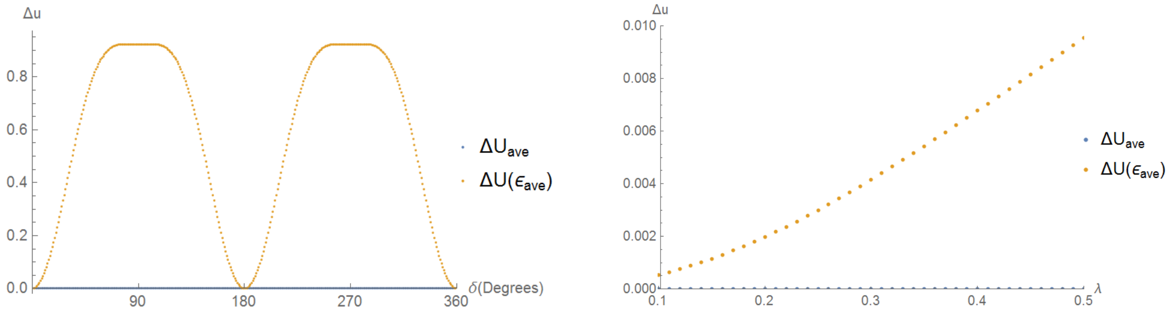
Figure 3.1: Δu_{ave} (Blue) and $\Delta u(\mathcal{E}_{ave})$ (orange) as a function of sequence length m. The amplitude damping and over rotation parameters of each of the 24 cliffords was randomly generated between [0,0.1]. Clearly, u_{fit} is better approximated by the average of the unitarities of each error (u_{ave}) than by the unitarity of the average error ($u(\boldsymbol{\mathcal{E}}_{ave})$).

(u_{ave}) is approximately equal to the unitarity obtained via XRB (u_{fit}). In addition, (u_{ave}) is a significantly better approximation of (u_{fit}) than the unitarity of the average error ($u(\boldsymbol{\mathcal{E}}_{ave})$) at all sequence lengths.

As seen in figure 3.2a, $u(\boldsymbol{\mathcal{E}}_{ave})$ has a strong dependence on the over-rotation angle. In the case where the error is unitary, the unitarity of each error is 1, so when we take the average over the set of gates we get $u_{ave} = 1$. Therefore u_{ave} can distinguish between coherent and incoherent errors. When averaging over the coherent over-rotation error, $\mathcal{E}_{ave}$ is a depolarizing channel which is incoherent; therefore $u(\mathcal{E}_{ave})$ cannot differentiate between coherent and incoherent errors, which causes it to have a dependence on the over-rotation parameter. This makes it a poor measurement of the unitarity of a gateset. For gates undergoing amplitude damping (figure 3.2b), $u(\mathcal{E}_{ave})$ again is much closer to u_{fit}, as we expect from looking at eq. (3.1).

Conclusion

For a single qubit system with gate-dependent amplitude dampening and/or over-rotation error, the unitarity obtained via XRB is the average of the unitarities of each gate. The average of the unitarity of each gate is able to differentiate between gate-dependent coherent



(a) Δu vs Over-Rotation

(b) Δu vs Amplitude Damping

Figure 3.2: Δu_{ave} (Blue) and $\Delta u(\mathcal{E}_{ave})$ (orange) calculated with respect to the over rotation angle α (a) and the amplitude damping (b). The average of a set of unitary gates is a depolarizing channel, which has a unitarity less than 1. This causes the large fluctuation of $\Delta u(\mathcal{E}_{ave})$ in (a). In (b), u_{ave} is a better approximation of u_{fit} than $u(\mathcal{E}_{ave})$ for decoherent gate-dependent noise.

errors (e.g., due to calibration errors) and incoherent errors (e.g., due to intrinsic noise processes). Knowledge about whether noise is coherent or incoherent in a system will have a significant impact on how experiments are constructed and implemented. In summary, we have shown how to interpret the output of the XRB protocol, and have proven analytically that under any gate-dependent error, the unitarity obtained via the XRB protocol can differentiate between coherent and decoherent gate-dependent error.

Chapter 4

From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters

4.1 Overview

Randomized benchmarking (RB) protocols have become an essential tool for providing a meaningful partial characterization of experimental quantum operations. While the RB decay rate is known to enable estimates of the average fidelity of those operations under gate-independent Markovian noise, under gate-dependent noise this rate is more difficult to interpret rigorously. In this paper, we prove that the single-qubit RB decay parameter p coincides with the decay parameter of the *gate-set circuit fidelity*, a novel figure of merit which characterizes the expected average fidelity over arbitrary circuits of operations from the gate-set. We also prove that, in the limit of high-fidelity single-qubit experiments, the possible alarming disconnect between the average gate fidelity and RB experimental results is simply explained by a basis mismatch between the gates and the state-preparation and measurement procedures, that is, to a unitary degree of freedom in labeling the Pauli matrices. Based on numerical evidence and physically motivated arguments, we conjecture that these results also hold for higher dimensions.

4.2 Introduction

The operational richness of quantum mechanics hints at an unprecedented computational power. However, this very richness carries over to a vast range of possible quantum error processes for which a full characterization is impractical for even a handful of quantum bits (qubits). Randomized benchmarking (RB) experiments [49, 99, 92, 46, 106, 107, 32, 43] were introduced to provide a robust, efficient, scalable, SPAM-independent[1], partial characterization of specific sets of quantum operations of interest, referred to as gate-sets. Such experiments have been widely adopted across all platforms for quantum computing, eg. [62, 42, 86, 10, 35, 156, 152, 111, 112], and have become a critical tool for characterizing and improving the design and control of quantum bits (qubits).

Recently it has been shown that RB experiments on an arbitrarily large number of qubits will always produce an exponential decay under arbitrary Markovian error models (that is, where errors are represented as completely-positive maps). This ensures a well-defined theoretical characterization of these experiments and enables an important test for the presence of non-Markovian errors, in spite of the gauge freedom between the experimental quantities and a theoretical figure of merit such as the average gate fidelity [131, 167, 116]. However, this theoretical advance still lacks a clear physical interpretation that rigorously connects the experimentally observed decay to a fidelity-based characterization of a physical set of gate- dependent errors. Linking an experimentally measured quantity to a physically meaningful figure of merit is not a mere intellectual satisfaction. It is crucial to ensure that a quantity measured in the context of a process characterization protocol indeed yields an outcome that assesses the quality of operations. What if a very noisy quantum device could yield a decent RB parameter? What if there exist scenarios where RB substantially underestimates the quality of a quantum device?

In this chapter, we show that in the regime of high fidelity gates on single qubits, a simple physical interpretation of RB data does exist and allows a reliable characterization of quantum operations. Further we conjecture, based on numerical evidence, that such an interpretation extends to arbitrary dimensions. Consequently, this work provides the theoretical foundation behind a fundamental tool for identifying and eliminating errors through examining the results of RB experiments.

Consider a targeted ideal gate-set $\mathbb{G} = \{\mathcal{G}\}$ and its noisy implementation $\tilde{\mathbb{G}} = \{\tilde{\mathcal{G}}\}$. We denote a circuit composed of m elements by

$$\tilde{\mathcal{G}}_{m:1} := \tilde{\mathcal{G}}_m \cdots \tilde{\mathcal{G}}_2 \tilde{\mathcal{G}}_1 \ . \tag{4.1}$$

---

[1] SPAM stands for "State preparation and measurement".

For leakage-free RB experiments with arbitrarily gate-dependent (but still Markovian) errors, the average probability of an outcome μ after preparing a state ρ and applying a circuit of $m+1$ operations that multiply to the identity is [167, 116]

$$\mathbb{E}_{\mathcal{G}_{m+1:1}} \left\langle \mu, \tilde{\mathcal{G}}_{m+1:1}(\rho) \right\rangle = Ap^m + B + \epsilon(m), \tag{4.2}$$

where $\langle M_1, M_2 \rangle := \mathrm{tr}\left\{ M_1^\dagger M_2 \right\}$ refers to the Hilbert-Schmidt inner product. On the right-hand side of eq. (4.2), A and B are independent of m (i.e., they only depend upon ρ, μ and $\tilde{\mathcal{G}}$) and $\epsilon(m)$ is a perturbative term that decays exponentially in m (decays significantly faster than p).

By design, RB gives some information about the error rate of motion-reversal (i.e., identity) circuits composed of gate-set elements. In this chapter, we show how this information relates to general circuits. Consider the traditional notion of *average fidelity* for a noisy circuit $\tilde{\mathcal{C}}$ to a target unitary circuit $\mathcal{C}$,

$$F(\tilde{\mathcal{C}}, \mathcal{C}) := \int \left\langle \tilde{\mathcal{C}}(\psi), \mathcal{C}(\psi) \right\rangle d\psi, \tag{4.3}$$

where the integral is taken uniformly over all pure states. Equation (4.3) corresponds to the definition of the usual notion of *average gate fidelity*, but is instead formulated in terms of "circuit", which is to be understood as a sequence of elementary gates. We introduce this nuance to define a novel figure of merit, the *gate-set circuit fidelity*, which compares all possible sequences of m implemented operations from the gate-set $\tilde{\mathbb{G}}$ to their targets in $\mathbb{G}$,

Definition 23 (*gate-set circuit fidelity*)**.**

$$\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m) := \mathbb{E}\left[F(\tilde{\mathcal{G}}_{m:1}, \mathcal{G}_{m:1}) \right] \ . \tag{4.4}$$

The case $m = 1$ yields the average fidelity of the gate-set $\tilde{\mathbb{G}}$ to $\mathbb{G}$. In general, the overall action of ideal circuits $\mathbb{G}_{m:1}$ is reproduced by $\tilde{\mathbb{G}}_{m:1}$ with fidelity $\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m)$. Having access to the gate-set circuit fidelity enables going beyond quantifying the quality of gate-set elements as it also quantifies the quality of circuits based on those elements. In this paper, we prove that for all single-qubit gate-sets with fidelities close to 1 and for an appropriately chosen targeted gate-set $\mathbb{G}$, the gate-set circuit fidelity can be closely estimated via RB experiments, for all circuit lengths m, even in cases of highly gate-dependent noise models. This is possible because it turns out that $\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m)$ essentially behaves like an exponential decay in m, uniquely parametrized by the RB decay constant p. The robust inclusion of gate-dependence is a major step forward since it encompasses very realistic noise models. We conjecture this result to hold for higher dimensions, based on numerical evidences and physically motivated arguments. Notice that the gate-set circuit fidelity quantifies the expected fidelity of *all* circuits (built from gate-set elements), and not only motion-reversal ones. This is an important observation to keep in mind because although RB experiments intrinsically revolve around motion-reversal circuits, the figure of merit that it yields isn't limited to such paradigm. Quantifying the quality of all circuits is much more useful than quantifying identity ones.

4.3 The dynamics of the gate-set circuit fidelity

It follows from the RB literature [49, 106] that for gate-independent noise models of the form $\tilde{\mathbb{G}} = \mathcal{E}\mathbb{G}$ or $\tilde{\mathbb{G}} = \mathbb{G}\mathcal{E}$, where $\mathcal{E}$ is a fixed error, the gate-set circuit fidelity behaves exactly as

$$\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m) = \frac{1}{d} + \frac{d-1}{d} p^m \ , \tag{4.5}$$

where p is estimated through standard RB by fitting to eq. (4.2) with $\epsilon(m) = 0$ and d is the dimension of the system. The relationship between the survival probability (p) decay curve and the decay in eq. (4.5) shouldn't be surprising. Indeed, consider a RB experiment with a noise model of the form $\mathcal{E}\mathbb{G}$ and a perfect inversion step $\mathcal{G}_{m+1} \in \mathbb{G}$ and perfect SPAM. In such a case, the gate-set circuit fidelity and the survival probability exactly coincide. A less trivial matter is to show the link between the RB decay parameter and eq. (4.5) for gate-dependent leakage-free noise models for which the choice of targeted gate-set is to be treated more carefully. In fact, as we will show, a poor choice of targeted gate-set can lead to a strong violation of eq. (4.5) in the sense that $1 - \mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m)$ can differ from $1 - (\frac{1}{d} + \frac{d-1}{d} p^m)$ by several orders of magnitude. An appropriate choice of targeted gate-set will essentially restore the decay relation shown in eq. (4.5).

Equation (4.5) generalizes to

$$\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, m) = \frac{1}{d} + \frac{d-1}{d} f_{tr}(\tilde{\mathbb{G}}, \mathbb{G}, m) \ , \tag{4.6}$$

where the fidelity on the traceless hyperplane is similar to the gate-set circuit fidelity, but is averaged over the traceless part of the pure states, $\psi_{\mathrm{tr}} = \psi - \ \mathbb{I}/d$:

$$f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, m) := \frac{\mathbb{E}\left(\int \left\langle \tilde{\mathcal{G}}_{m:1}(\psi_{\mathrm{tr}}), \mathcal{G}_{m:1}(\psi_{\mathrm{tr}}) \right\rangle d\psi \right)}{\int \langle \psi_{\mathrm{tr}}, \psi_{\mathrm{tr}} \rangle d\psi} \ . \tag{4.7}$$

The integrand in the numerator of the right-hand side of eq. (4.7) can be visualized as the fidelity restricted on the Bloch space, comparing the ideally mapped Bloch vectors $\psi_{\mathrm{tr}} \to \mathcal{G}_{m:1}(\psi_{\mathrm{tr}})$ to their noisy analog $\tilde{\mathcal{G}}_{m:1}(\psi_{\mathrm{tr}})$. Equation (4.6) is quickly obtained by realizing that the symmetric integral over the Bloch space evaluates to zero.

Under gate-dependent noise, $1 - f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, 1)$ could differ from $1 - p$ by several orders of magnitude [131, 135]. Such discrepancy was seen as a serious concern: the observed RB decay seemingly fails in characterizing the quality of quantum operations. To see the possible immense disconnect between p and $f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, 1)$, consider the canonical example where single-qubit gates are perfectly implemented, but differ from the targets $\mathcal{G} \in \mathbb{G}$ by a labeling of the Pauli axes:

$$\tilde{\mathcal{G}}(X) = \mathcal{G}(Y) \ , \tag{4.8a}$$

$$\tilde{\mathcal{G}}(Y) = \mathcal{G}(Z) \ , \tag{4.8b}$$

$$\tilde{\mathcal{G}}(Z) = \mathcal{G}(X) \ . \tag{4.8c}$$

This noise model would lead to an absence of decay in the survival probability, that is $p = 1$. Indeed, motion-reversal circuits are perfectly implementing the identity gate, regardless of the length of the circuit. A quick calculation results in $f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, m) = 0$, which translates to a difference in orders of magnitude $|\log(1-p) - \log\Big(1 - f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, 1)\Big)|$ that tends to infinity as $p \to 1$. The RB experiment indicates no operational error while the average gate fidelity is $1/2$. Does the outcome of RB massively underestimate the error? Notice that since the implementation error is a permutation of labels, there is actually no observable error in the device. The alarmingly low value of gate-set circuit fidelity of $\tilde{\mathbb{G}}$ to $\mathbb{G}$ is simply a consequence of a poor choice of targeted gate-set.

As a more involved example, let the noise model be $\tilde{\mathbb{G}} = \mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ for any non-identity unitary channel $\mathcal{U}$ and let the set of targeted operations be $\mathbb{G}$ (this includes our previous mislabeling example as a special scenario). In such cases $f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, 1)$ can take any value in the interval $[0, 1)$, depending on the choice of $\mathcal{U}$. However, using the same argument as in the previous example, the survival probability is not subject to a decay ($p = 1$), showing once again how the decay parameter could arbitrarily differ from a poorly defined average gate fidelity. This apparent disconnect arises due to a *basis mismatch* between the bases in which the noisy gate-set and the targeted gate-set are defined. A reconciliation of the RB observations with a gate-set circuit fidelity is obtained by changing the set of targets to $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ since $f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, 1) = 1$. One might argue that implementing $\tilde{\mathbb{G}} = \mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ instead of the ideal $\mathbb{G}$ should raise an operational error. Not necessarily: consider a circuit uniquely constructed from operations $\tilde{\mathcal{G}}_i \in \tilde{\mathbb{G}}$. According to Born's rule, the probability of measuring the outcome i associated with the positive operator μ_i after performing the circuit on a state ρ is:

$$\begin{aligned}
p_i &= \left\langle \mu_i, \tilde{\mathcal{G}}_{m:1}(\rho) \right\rangle \\
&= \left\langle \mu_i, \mathcal{U}\mathcal{G}_m\mathcal{U}^\dagger \cdots \mathcal{U}\mathcal{G}_2\mathcal{U}^\dagger\mathcal{U}\mathcal{G}_1\mathcal{U}^\dagger(\rho) \right\rangle \\
&= \left\langle \mu_i, \mathcal{U}\mathcal{G}_{m:1}\mathcal{U}^\dagger(\rho) \right\rangle \\
&= \left\langle \mu_i', \mathcal{G}_{m:1}(\rho') \right\rangle \ , \qquad (4.9)
\end{aligned}$$

where $\rho' = \mathcal{U}^\dagger(\rho)$, $\mu_i' = \mathcal{U}^\dagger(\mu_i)$. That is, the error can be interpreted as part of SPAM procedures instead of operations. Since the unitary transformation can be pushed to either SPAM procedures or coherent manipulations, it should be seen as a mismatch between them. Indeed, the physical unitary conjugation $\tilde{\mathbb{G}} = \mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ doesn't affect the *internal action* of operations, but exclusively the connection between operations and SPAM procedures. Changing the targeted gate-set $\mathbb{G}$ to $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ is allowed by the degree of freedom in labeling the basis for SPAM procedures and the basis for processes.

In section 4.6.1, we show how exactly the disconnect between p^m and $f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m)$ depends on the choice of targeted gate-set $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$. That is, we provide an expression of the form

$$f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m) = C(\mathcal{U})p^m + D(m, \mathcal{U}) \ , \qquad (4.10)$$

where $\mathcal{U}$ is a physical unitary channel (see theorem 25). A first interesting observation is that $D(m, \mathcal{U})$ is typically negligible or becomes rapidly negligible as it is also exponentially suppressed in m[2]. This means that the relative variation in f_{tr} as the circuit grows in length,

$$\frac{f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m+1)}{f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m)} = p + \delta(m, \mathcal{U}) \ , \qquad (4.11)$$

depends weakly on the choice of targeted gate-set. More precisely, $\delta(m, \mathcal{U})$ is composed of two factors: the first one decays exponentially in m and is at most of order $(1-p)^{m/2}$, while the second carries the dependence in $\mathcal{U}$; the existence of a specific choice of $\mathcal{U}$ such that this last factor becomes at most of order $(1-p)^{3/2}$ is proven in the single-qubit case (section 4.6.2), and conjectured to hold in general. The explicit behaviour of $\delta(m, \mathcal{U})$ given a numerically simulated gate-dependent noise model is illustrated in fig. 4.1.

---

[2] Since $D(1, \mathcal{U})$ is typically close to 0, the exponential suppression is quite effective compared to $p^m \approx 1 - m(1-p)$ which is essentially linear for small m.

Consequently, the gate-set circuit fidelity can be updated with a good approximation through the recursion relation

$$\mathcal{F}(\tilde{\mathbb{G}},\mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m+1) \approx \frac{1}{d} + p\left(\mathcal{F}(\tilde{\mathbb{G}},\mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m) - \frac{1}{d}\right). \tag{4.12}$$

Roughly speaking, this means that the choice of basis $\mathcal{U}$ selected to express the targets, $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$, is not highly significant when it comes to updating the gate-set circuit fidelity as the circuit grows in depth. The RB decay rate p enables the decrease in fidelity due to adding a gate to a circuit to be predicted.

However, to provide insight on the total value of the gate-set circuit fidelity given a circuit's length m, we need a stronger relation between the RB estimate of p and the gate-set circuit fidelity. Fortunately, the freedom in the choice of targeted gate-set allows us to choose a basis which will facilitate estimation of the total change in gate-set circuit fidelity for arbitrary circuit lengths.

In section 4.6.2, we prove that the potentially large disconnect between p and $f_{\mathrm{tr}}(\tilde{\mathbb{G}},\mathcal{U}\mathbb{G}\mathcal{U}^\dagger, 1)$ under general gate-dependent noise is almost completely accounted for by a basis mismatch which, as we argued earlier, does not exactly correspond to a process error since unitary conjugation does not affect the internal dynamics of operations.

Proposition 1. *For any single-qubit noisy gate-set* $\tilde{\mathbb{G}}$ *perturbed from* $\mathbb{G}$*, there exists an ideal targeted gate-set* $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$*, where* $\mathcal{U}$ *is a physical unitary, such that*

$$\mathcal{F}(\tilde{\mathbb{G}},\mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m) = \frac{1}{d} + \frac{d-1}{d}p^m + O\left((1-p)^2\right) . \tag{4.13}$$

In fact, we conjecture this result to hold for any dimension, or at least for most realistic gate-dependent noise models. To grasp the physical reasoning behind this, we refer to the end of section 4.6.2, as it rests on some prior technical analysis. The extension of proposition 1 to 2-qubit systems is supported by numerical evidence (see sections 4.6.1 to 4.6.2).

The unitary freedom appearing in the gate-set circuit fidelity means that there exist infinitely many fidelity-based figures of merit describing noisy circuits, one for each targeted gate-set $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$. Of course, there exist choices of targeted operations that yield gate-set circuit fidelities that differ from eq. (4.13) (see [131, 135]); the example shown in eqs. (4.8a) to (4.8c) is an elementary instance thereof. Proposition 1 simply states that there exists a natural choice of gate-set $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ that connects the outcome of an RB experiment to a gate-set circuit fidelity. The choice of basis $\mathcal{U}$ allows us to take either the perspective of the gates or the perspective of SPAM procedures. This is implicitly done when defining gates relative to the energy eigenbasis of the system. In this picture, the gate-set circuit fidelity describes the accuracy of the internal behaviour of operations as they act in concert.

To reformulate the result, the family of circuits $\tilde{\mathcal{G}}_{m:1}$ built from a composition of m noisy operations $\tilde{\mathcal{G}} \in \tilde{\mathbb{G}}$ mimics the family of ideal circuits $\mathcal{U}\mathcal{G}_{m:1}\mathcal{U}^\dagger$ with fidelity $\frac{1}{d}+\frac{d-1}{d}p^m$. In the paradigm where the initially targeted operations $\mathcal{G} \in \mathbb{G}$ are defined with respect to SPAM procedures, $\mathcal{U}$ captures the misalignment between the basis in which the operations $\tilde{\mathcal{G}} \in \tilde{\mathbb{G}}$ are defined and the basis defined by SPAM procedures. This goes farther: consider an additional gate-set $\mathbb{H}$, for which the targeted operations $\mathcal{H} \in \mathbb{H}$ are also defined with respect to SPAM procedures. From proposition 1, there exists a physical unitary $\mathcal{V}$ for which $\tilde{\mathcal{H}}_{m:1}$ imitates the action of $\mathcal{V}\mathcal{H}_{m:1}\mathcal{V}^\dagger$ with fidelity $\frac{1}{d}+\frac{d-1}{d}q^m$ (where q is estimated through RB). $\mathcal{U}^\dagger\mathcal{V}$ captures the basis mismatch between the gate-sets $\tilde{\mathbb{G}}$ and $\tilde{\mathbb{H}}$. A non-trivial mismatch could easily be imagined if, for instance, gates belonging to $\tilde{\mathbb{H}}$ were obtained through a different physical process than $\tilde{\mathbb{G}}$, or calibrated with regards to alternate points of reference.

4.4 Finding the appropriate set of targeted gates for specific noise models

We now discuss how the appropriate unitary conjugation on the initial targeted gate-set can be calculated for specific noise models, whether from numerical simulations, analytic approximations, or tomographic reconstructions. As shown in theorem 25 and eq. (4.10), the total change of gate-set circuit fidelity depends on the physical basis in which the ideal gate-set is expressed. In the single-qubit case, we showed the existence of a physical basis $\mathcal{U}$ that reconciles $f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m)$ with p^m through proposition 1. One might suspect that the unitary $\mathcal{U}$ can be found through the maximization of the gate-set fidelity:

$$\mathcal{U} = \underset{\mathcal{V}}{\operatorname{argmax}}\, \mathcal{F}(\tilde{\mathbb{G}}, \mathcal{V}\mathbb{G}\mathcal{V}^\dagger, 1)\ , \tag{4.14}$$

and indeed this would handle noise models of the form $\tilde{\mathbb{G}} = \mathcal{U}\mathcal{E}\mathbb{G}\mathcal{U}^\dagger$, as

$$p = f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, 1) \geq f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, 1)\ .$$

However, this hypothesis fails for simple noise models of the form $\tilde{\mathbb{G}} = \mathcal{U}^\dagger\mathcal{E}\mathbb{G}\mathcal{U}^\dagger$, where

$$p = f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, 1) \leq f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathbb{G}, 1)\ .$$

These examples show that p can be greater or less than $f_{\mathrm{tr}}(\tilde{\mathbb{G}},\mathbb{G},1)$, depending on the noise model. More examples are derived in [131, 135]. This particular case study is informative as these two last noise models share something in common: there exists a choice of unitary $\mathcal{U}$ that cancels the noisy map on the right of the noisy gate-set. Although such exact cancellation is not always possible, we now show that a close approximation is sufficient. Consider the slightly more general noise model of the form $\tilde{\mathbb{G}} = \mathcal{E}_L\mathbb{G}\mathcal{E}_R$, where we allow fixed but arbitrary error maps to the left and the right of an ideal gate-set. It can be shown while staying under the scope of the original analysis provided in [106, 107] that $p^m = f_{\mathrm{tr}}(\mathcal{E}_R\mathcal{E}_L\mathbb{G},\mathbb{G},m)$, since $\mathcal{E}_R\mathcal{E}_L$ is the effective error map between two otherwise perfect implementations of the gate-set elements. In the single-qubit case (and for many, if not all physically motivated higher dimensional noise models) there exists a unitary operation $\mathcal{U}$ such that

$$F(\mathcal{E}_R\mathcal{E}_L,\ \mathbb{I}) = \mathcal{F}(\mathcal{E}_L\mathbb{G}\mathcal{E}_R,\ \mathcal{U}\mathbb{G}\mathcal{U}^\dagger,1) + O((1-p)^2)\ , \tag{4.15}$$

(see section 4.6.2). That is, the fidelity of the map between two noisy gate-sets can be seen as the gate-set circuit fidelity between a noisy gate-set and an appropriately targeted ideal one. A choice of such physical unitary is

$$\mathcal{U} = \underset{\mathcal{V}}{\operatorname{argmax}}\, F\left(\mathcal{E}_R\mathcal{V},\mathbb{I}\right)\ , \tag{4.16}$$

which essentially cancels the unitary part of $\mathcal{E}_R$ [3]. Another way to see this is that the unitary freedom allows us to reexpress the errors $\mathcal{E}_L,\ \mathcal{E}_R$ as

$$\begin{aligned}\mathcal{E}_L &\rightarrow \mathcal{U}^\dagger\mathcal{E}_L\\ \mathcal{E}_R &\rightarrow \mathcal{E}_R\mathcal{U}\ .\end{aligned}$$

We can then choose the unitary that depletes $\mathcal{E}_R\mathcal{U}$ from any coherent component. Intuitively, reexpressing the error on one side to make it incoherent prevents any type of unitary conjugation of the form $\tilde{\mathbb{G}} = \mathcal{U}\mathcal{E}\mathbb{G}\mathcal{U}^\dagger$.

For more general gate-dependent noise models, the idea remains more or less the same. As shown in section 4.6.2, the right error $\mathcal{E}_R$ is replaced by its generalization, the 4th order right error $\mathcal{E}_R^{(4)} = \mathbb{E}\left[\mathcal{G}_{4:1}{}^\dagger\tilde{\mathcal{G}}_{4:1}\right]$ (eq. (4.38a)). From there, we find:

Proposition 2 (Finding the appropriate targeted gate-set)**.** *A proper choice of physical basis $\mathcal{U}$ for which eq. (4.13) applies is*

$$\mathcal{U} = \underset{\mathcal{V}}{\operatorname{argmax}}\, F\left(\mathbb{E}\left[\mathcal{G}_{4:1}{}^\dagger\tilde{\mathcal{G}}_{4:1}\right]\mathcal{V},\mathbb{I}\right)\ . \tag{4.17}$$

---

[3] Of course, $\underset{\mathcal{V}}{\operatorname{argmax}}\, F\left(\mathcal{V}^\dagger\mathcal{E}_L,\mathbb{I}\right)$ would also fulfill eq. (4.15).

For this choice, $\mathcal{U}$ cancels the unitary part of the 4^{th} *order right error.*

This provides a means to guide the search for the appropriate choice of ideal targeted gate-set of comparison $\mathcal{U}\mathbb{G}\mathcal{U}^{\dagger}$ given a numerical noise model $\tilde{\mathbb{G}}$. Indeed, the 4^{th} order right error is easily found, either by direct computation of the average $\mathbb{E}\left[\mathcal{G}_{4:1}{}^{\dagger}\tilde{\mathcal{G}}_{4:1}\right]$, or more efficiently by solving the eigensystem defined in eq. (4.25a). The optimization defined in eq. (4.17) can be solved via a gradient ascent parametrized over the d^2-1 degrees of freedom of $SU(d)$.

In the single-qubit case, the optimization procedure can be replaced by an analytical search. Given the process matrix $\boldsymbol{\mathcal{E}}_R^{(4)}$ of the 4^{th} order right error, it suffices to find the polar decomposition of its 3×3 submatrix acting on the Bloch vectors: $\boldsymbol{\mathcal{E}}_R^{(4)}\boldsymbol{\Pi}_{\text{tr}} = \boldsymbol{\mathcal{D}}_{\text{tr}}\boldsymbol{\mathcal{V}}_{\text{tr}}$. The unitary factor $\mathcal{V}$ corresponds to $\mathcal{U}^{\dagger}$, while the positive factor $\mathcal{D}$ captures an incoherent process (rigorously defined in eq. (4.41)).

With this at hand, we performed numerically simulated RB experiments under gate-dependent noise models. Each of the 24 Cliffords was constructed by a sequence of X and Y pulses, $G_x = P(X,\pi/2)$ and $G_y = P(Y,\pi/2)$, where

$$P(H,\theta) := e^{i\theta H/2}\ . \tag{4.18}$$

The 2-qubit Cliffords were obtained through the construction shown in [10, 42], where the 11520 gates are composed of single-qubit Clifford and CZ gates. The implementation of the 2-qubit entangling operation was consistently performed with an over-rotation: $\tilde{\mathcal{G}}_{CZ} = P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1})$. In fig. 4.2, the single-qubit gate generators are modeled with a slight over-rotation: $\tilde{\mathcal{G}}_x = P(X,\pi/2+10^{-1})$ and $\tilde{\mathcal{G}}_y = P(Y,\pi/2+10^{-1})$. This model exemplifies the failure of the maximization hypothesis proposed in eq. (4.14). In figs. 4.1 and 4.3, the single-qubit gate generators are followed by a short Z pulse, $\tilde{\mathcal{G}}_x = P(\sigma_z,\theta_z)\mathcal{G}_x$ and $\tilde{\mathcal{G}}_y = P(\sigma_z,\theta_z)\mathcal{G}_y$, which reproduces the toy model used in [131].

4.5 Conclusion

RB experiments estimate the survival probability decay parameter p of motion-reversal circuits constituted of operations from a noisy gate-set $\tilde{\mathbb{G}}$ of increasing length (see eq. (4.2)). While motion-reversal is intrinsic to the experimental RB procedure, the estimated decay constant p can be interpreted beyond this paradigm. In this chapter we have shown that, in a physically relevant limit, the very same parameter determines an interesting figure of merit, namely the gate-set circuit fidelity (defined in eq. (4.4)): as a random operation from $\tilde{\mathbb{G}}$ is introduced to a random circuit constructed from elements in $\tilde{\mathbb{G}}$, p captures the expected relative change in the gate-set circuit fideilty through eq. (4.12).

It is also possible to characterize the full evolution of gate-set circuit fidelity as a function of the circuit length. In this chapter, we have also demonstrated that given a single-qubit noisy gate-set $\tilde{\mathbb{G}}$ perturbed from $\mathbb{G}$, there exists an alternate set of targeted gates obtained through a physical basis change $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ such that the gate-set circuit fidelity takes the simple form given in eq. (4.13). This gives a rigorous underpinning to previous work that has assumed that the experimental RB decay parameter robustly determines a relevant average gate fidelity (eq. (4.3)) for experimental control under generic gate-dependent scenarios. We conjecture a similar result to hold for higher dimensions and provide numerical evidence and physically motivated arguments to support this conjecture. Given any specific numerical noise model $\tilde{\mathbb{G}}$ perturbed from $\mathbb{G}$, we showed how to obtain a physical unitary $\mathcal{U}$ for which eq. (4.13) holds. The procedure can be seen as a fidelity maximation of the 4th order right error acting on the gate-set through a unitary correction (see proposition 2).

The introduction of such a physical basis adjustment is natural because it has no effect on how errors accumulate as a function of the sequence length. Rather, it only reflects a basis mismatch to the experimental SPAM procedures. This is in principle detectable by RB experiments but in practice not part of the goals of such diagnostic experiments. In particular, differences in the (independent) basis adjustments required for distinct gate-sets will not appear in any characterization of the individual gate-sets, but will be detected when comparing RB experiments for this distinct gate-sets (e.g., comparing dihedral benchmarking and standard randomized benchmarking experiments which have distinct gate-sets but share gates in common, or comparing independent single-qubit RB on two qubits - which has no two-qubit entangling gate - with standard two-qubit RB). We leave the problem of characterizing relative basis mismatch between independent gate-sets as a subject for further work.

4.6 Supplementary Material

4.6.1 An expression for the total change in the gate-set circuit fidelity

In this section, we extend the standard RB analysis under gate-dependent noise provided in [167, 116] in order to prove the claim from eq. (4.11) that standard RB returns the relative variation of the gate-set circuit fidelity.

Let $\boldsymbol{\mathcal{A}}$ be the Liouville matrix of a linear map $\mathcal{A}$ and $\Pi_{\rm tr}(\rho) = \rho - \mathbb{I}\,\mathrm{tr}\,\rho/d$ be the projector onto the traceless component. We denote the Frobenius norm, which is defined by the Hilbert-Schmidt inner product, as $\|\cdot\|_F$. For instance, in the qubit case $\|\boldsymbol{\Pi}_{\rm tr}\|_F^2 = 3$. We denote the induced 2-norm as $\|\cdot\|_2$, which corresponds to the maximal singular value. Let e_j be the canonical unit vectors, $A = \sum_{j,k} a_{j,k} e_j e_k^T$, and

$$\mathrm{vec}(A) = \sum_{j,k} a_{j,k} e_k \otimes e_j \ . \tag{4.19}$$

Using the identity

$$\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B) \ , \tag{4.20}$$

we have

$$\begin{aligned} f_{\rm tr}(\tilde{\mathbb{G}}, \mathbb{G}, m) &= \mathbb{E}\left(\frac{\left\langle \tilde{\boldsymbol{\mathcal{G}}}_{m:1}\boldsymbol{\Pi}_{\rm tr}, \boldsymbol{\mathcal{G}}_{m:1}\boldsymbol{\Pi}_{\rm tr}\right\rangle}{\|\boldsymbol{\Pi}_{\rm tr}\|_F^2}\right) \\ &= \frac{\mathrm{vec}^\dagger(\boldsymbol{\Pi}_{\rm tr})}{\|\boldsymbol{\Pi}_{\rm tr}\|_F} \boldsymbol{\mathcal{T}}^m \frac{\mathrm{vec}(\boldsymbol{\Pi}_{\rm tr})}{\|\boldsymbol{\Pi}_{\rm tr}\|_F} \end{aligned} \tag{4.21}$$

where the twirling superchannel [37, 131, 167] is

$$\boldsymbol{\mathcal{T}} = \mathbb{E}[\boldsymbol{\mathcal{G}}_{\rm tr} \otimes \tilde{\boldsymbol{\mathcal{G}}}] \tag{4.22}$$

and $\boldsymbol{\mathcal{G}}_{\rm tr} = \boldsymbol{\mathcal{G}}\boldsymbol{\Pi}_{\rm tr}$. Changing the gate-set $\mathbb{G}$ to $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ for some physical unitary $\mathcal{U}$ leaves $\boldsymbol{\Pi}_{\rm tr} = \boldsymbol{\mathcal{U}}\boldsymbol{\Pi}_{\rm tr}\boldsymbol{\mathcal{U}}^\dagger$. Therefore

$$f_{\rm tr}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m) = \frac{\mathrm{vec}^\dagger(\boldsymbol{\mathcal{U}}\boldsymbol{\Pi}_{\rm tr})}{\|\boldsymbol{\Pi}_{\rm tr}\|_F} \boldsymbol{\mathcal{T}}^m \frac{\mathrm{vec}(\boldsymbol{\mathcal{U}}\boldsymbol{\Pi}_{\rm tr})}{\|\boldsymbol{\Pi}_{\rm tr}\|_F} \ . \tag{4.23}$$

The spectrum of $\boldsymbol{\mathcal{T}}$ is unchanged under the basis change $\boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{U}}^\dagger$. Moreover, its most important eigenvectors are as follows:

Lemma 24. *Let p be the highest eigenvalue of $\boldsymbol{\mathcal{T}}$ and*

$$\boldsymbol{\mathcal{A}}_m := p^{-m}\mathbb{E}\left[(\boldsymbol{\mathcal{G}}_{{\rm tr},m:1})^\dagger \boldsymbol{\Pi}_{\rm tr} \tilde{\boldsymbol{\mathcal{G}}}_{m:1}\right] \ , \tag{4.24a}$$

$$\boldsymbol{\mathcal{B}}_m := p^{-m}\mathbb{E}\left[\tilde{\boldsymbol{\mathcal{G}}}_{m:1} \boldsymbol{\Pi}_{\rm tr} (\boldsymbol{\mathcal{G}}_{{\rm tr},m:1})^\dagger\right] \ . \tag{4.24b}$$

Then we have

$$\mathrm{vec}^\dagger(\boldsymbol{\mathcal{A}}_\infty^T)\boldsymbol{\mathcal{T}} = p\ \mathrm{vec}^\dagger(\boldsymbol{\mathcal{A}}_\infty^T) \ , \tag{4.25a}$$

$$\boldsymbol{\mathcal{T}}\mathrm{vec}(\boldsymbol{\mathcal{B}}_\infty) = p\ \mathrm{vec}(\boldsymbol{\mathcal{B}}_\infty) \ . \tag{4.25b}$$

Proof. By eq. (4.20),

$$\mathrm{vec}(\boldsymbol{\mathcal{B}}_m) = p^{-m}\mathbb{E}((\boldsymbol{\mathcal{G}}_{\mathrm{tr},m:1})^* \otimes \tilde{\boldsymbol{\mathcal{G}}}_{m:1})\mathrm{vec}(\boldsymbol{\Pi}_{\mathrm{tr}}) \ . \tag{4.26}$$

As the Liouville representation is real-valued and the $\boldsymbol{\mathcal{G}}_j$ are independent,

$$\mathrm{vec}(\boldsymbol{\mathcal{B}}_m) = (\boldsymbol{\mathcal{T}}/p)^m\,\mathrm{vec}(\boldsymbol{\Pi}_{\mathrm{tr}}) \ . \tag{4.27}$$

Since the noisy gate-set $\tilde{\mathbb{G}}$ is a small perturbation from $\mathbb{G}$ the spectrum of $\boldsymbol{\mathcal{T}}$ will be slightly perturbed from $\{1,0,0,\ldots\}$. Therefore $(\boldsymbol{\mathcal{T}}/p)^m$ approaches a rank 1 projector as m increases and so $\mathrm{vec}(\boldsymbol{\mathcal{B}}_\infty)$ is a $+1$-eigenvector of $\boldsymbol{\mathcal{T}}/p$.

The same argument applies to $\boldsymbol{\mathcal{A}}_\infty^T$. □

Lemma 24 allows us to write

$$\boldsymbol{\mathcal{T}} = p\,\frac{\mathrm{vec}(\boldsymbol{\mathcal{B}}_\infty)\mathrm{vec}^\dagger(\boldsymbol{\mathcal{A}}_\infty^T)}{\langle \boldsymbol{\mathcal{A}}_\infty^T, \boldsymbol{\mathcal{B}}_\infty\rangle} + \boldsymbol{\Delta} \ , \tag{4.28}$$

with $\boldsymbol{\Delta}\mathrm{vec}\left(\boldsymbol{\mathcal{B}}_\infty\right) = \mathrm{vec}^\dagger\left(\boldsymbol{\mathcal{A}}_\infty^T\right)\boldsymbol{\Delta} = 0$. In eq. (4.23), we can expand the vectors as

$$\frac{\mathrm{vec}^\dagger(\boldsymbol{\mathcal{U}}\boldsymbol{\Pi}_{\mathrm{tr}})}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F} = a(\mathcal{U})\frac{\mathrm{vec}^\dagger(\boldsymbol{\mathcal{A}}_\infty^T)}{\|\boldsymbol{\mathcal{A}}_\infty\|_F} + \sqrt{1-a^2(\mathcal{U})}w^\dagger(\mathcal{U}) \tag{4.29a}$$

$$\frac{\mathrm{vec}(\boldsymbol{\mathcal{U}}\boldsymbol{\Pi}_{\mathrm{tr}})}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F} = b(\mathcal{U})\frac{\mathrm{vec}(\boldsymbol{\mathcal{B}}_\infty)}{\|\boldsymbol{\mathcal{B}}_\infty\|_F} + \sqrt{1-b^2(\mathcal{U})}v(\mathcal{U}) \tag{4.29b}$$

where

$$a(\mathcal{U}) := \frac{\langle \boldsymbol{\mathcal{A}}_\infty^T, \boldsymbol{\mathcal{U}}\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2}\left(\frac{\|\boldsymbol{\mathcal{A}}_\infty\|_F^2}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2}\right)^{-1/2} \ , \tag{4.30}$$

$$b(\mathcal{U}) := \frac{\langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{B}}_\infty\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2}\left(\frac{\|\boldsymbol{\mathcal{B}}_\infty\|_F^2}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2}\right)^{-1/2} \ . \tag{4.31}$$

and $v(\mathcal{U})$, $w(\mathcal{U})$ are implicitly defined unit vectors. Using this expansion together with eq. (4.28) in eq. (4.23) yields the following result:

Theorem 25 (Total gate-set circuit fidelity)**.** *The gate-set circuit fidelity obeys*

$$\mathcal{F}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m) = \frac{1}{d} + \frac{d-1}{d}\left(C(\mathcal{U})p^m + D(m,\mathcal{U})\right) \ , \tag{4.32}$$

where

$$\begin{aligned}C(\mathcal{U}) :=& \frac{\left\langle \boldsymbol{\mathcal{A}}_\infty^T, \boldsymbol{\mathcal{U}}\right\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} \frac{\langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{B}}_\infty\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} \left(\frac{\left\langle \boldsymbol{\mathcal{A}}_\infty^T, \boldsymbol{\mathcal{B}}_\infty\right\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} \right)^{-1} \\ =& \frac{\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{A}}_\infty \boldsymbol{\mathcal{U}}\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} \frac{\left\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{U}}^\dagger \boldsymbol{\mathcal{B}}_\infty \right\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} \frac{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2}{\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{A}}_\infty \boldsymbol{\mathcal{B}}_\infty\rangle} \qquad (4.33a) \\ D(m,\mathcal{U}) :=& \sqrt{1-a^2(\mathcal{U})}\sqrt{1-b^2(\mathcal{U})} w(\mathcal{U})^\dagger \boldsymbol{\Delta}^m v(\mathcal{U}) \ . \qquad (4.33b)\end{aligned}$$

In [131, 167, 116] it is shown that standard RB provides an estimate of p. Notice that p is independent of the basis in which the ideal gate-set of comparison, $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$, is expressed.

From eq. (4.32), it is straightforward to show that

$$\begin{aligned}\delta(m,\mathcal{U}) :=& \frac{f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m+1)}{f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m)} - p = \\ & \sqrt{1-a^2(\mathcal{U})}\sqrt{1-b^2(\mathcal{U})} \frac{w(\mathcal{U})^\dagger \boldsymbol{\Delta}^m (\boldsymbol{\Delta} - p\boldsymbol{\Pi}_{\mathrm{tr}}) v(\mathcal{U})}{f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{U}\mathbb{G}\mathcal{U}^\dagger, m)} \ , \qquad (4.34)\end{aligned}$$

which is exponentially suppressed. We show in the next section that the eigenvalues of $\boldsymbol{\Delta}$ are at most of order $\sqrt{1-p}$, which ensures a very fast decay, as shown in fig. 4.1. Equation (4.11) is in fact a reformulation of eq. (4.34).

4.6.2 Varying the ideal gate-set of comparison

In this section, we prove proposition 1 by determining how the basis $\mathcal{U}$ of the ideal gate-set $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$ affects the coefficients in eq. (4.32).

Let $\mathbb{G}$ be an ideal gate set defined with respect to the SPAM procedures. We can write the elements of a noisy gate-set as

$$\tilde{\boldsymbol{\mathcal{G}}} = \boldsymbol{\mathcal{G}} + \boldsymbol{\delta}_{\mathcal{G}}^{(I)} \boldsymbol{\mathcal{G}} \ , \qquad (4.35)$$

so that the perturbations $\delta_{\mathcal{G}}$ both capture the errors in the noisy gate and the mismatch with the targeted computational basis. Under gate-independent noise with no basis mismatch, $\tilde{\mathbb{G}} = \mathcal{E}\mathbb{G}$ and the infidelity of the perturbed operations $\mathbb{I} + \delta_{\mathcal{G}}^{(I)}$ is $r(\mathcal{E}) := 1 - F(\mathcal{E}, \mathbb{I})$. A basis mismatch will change the infidelity of the perturbations roughly to $r(\mathcal{U}\mathcal{E}) + r(\mathcal{U}^\dagger)$ for some unitary channel $\mathcal{U}$, which will typically differ substantially from the fidelity inferred from the associated RB experiment.

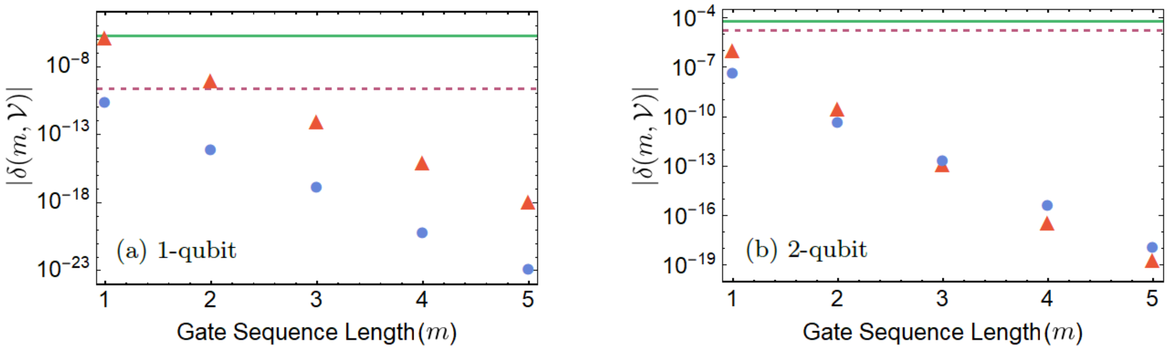


Figure 4.1: Absolute value of the deviation $\delta(m,\mathcal{V})$, described in eq. (4.11) (also see eq. (4.34)), as function of circuit length m with noise model generated by $\tilde{\mathcal{G}}_x = P(\sigma_z, 10^{-1})\mathcal{G}_x$ and $\tilde{\mathcal{G}}_y = P(\sigma_z, 10^{-1})\mathcal{G}_y$, $\tilde{\mathcal{G}}_{CZ} = P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1})$ (see eq. (2.30)). The red triangles are obtained with the choice of basis $\mathcal{V} = \mathbb{I}$, while the blue circles are obtained with the choice $\mathcal{V} = \mathcal{U}$ where $\mathcal{U}$ is found through eq. (4.17). The purple horizontal dashed line corresponds to $(1-p)^2$, while the full green line corresponds to $(1-\mathcal{F}(\tilde{\mathbb{G}}, \mathbb{G}, 1))^2$. For both ideal gate-sets $\mathbb{G}$ and $\mathcal{U}\mathbb{G}\mathcal{U}^\dagger$, the deviation becomes quickly negligible as the sequence length increases. In fact, in the case $\mathcal{V} = \mathcal{U}$ (blue circles), the deviation is always below $(1-p)^2$.

Experimentally, such basis mismatches will be relatively small as operations will be somewhat consistent with SPAM procedures. Under this assumption, we now show that there exists an alternate perturbative expansion,

$$\tilde{\boldsymbol{\mathcal{G}}} = \boldsymbol{\mathcal{U}\mathcal{G}\mathcal{U}}^\dagger + \boldsymbol{\delta}_{\mathcal{G}}^{(U)}\boldsymbol{\mathcal{U}\mathcal{G}\mathcal{U}}^\dagger \ , \tag{4.36}$$

for which $r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(U)})$ is in line with the data resulting from an RB experiment.

In section 4.6.1, we showed that $(\boldsymbol{\mathcal{T}}/p)^n$ converges to a rank-1 projector. We now quantify the rate of convergence. Recall that $\boldsymbol{\mathcal{T}}$ is perturbed from a rank-1 projector with spectrum $\{1, 0, 0, \cdots\}$. Hence, by the Bauer-Fike theorem [14], for any eigenvalue $\lambda \neq p$

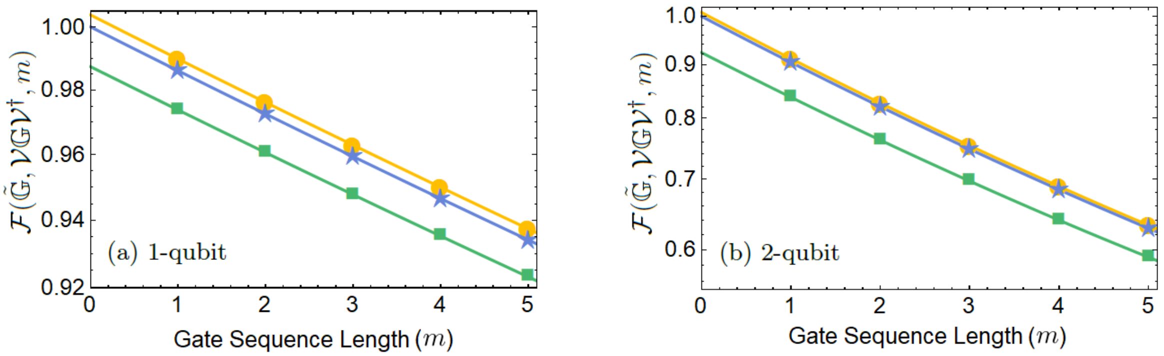


Figure 4.2: gate-set circuit fidelity $\mathcal{F}(\tilde{\mathbb{G}}, \mathcal{V}\mathbb{G}\mathcal{V}^\dagger, m)$ as a function of circuit length m with noise model generated by $\tilde{\mathcal{G}}_x = P(X, \pi/2 + 10^{-1})$, $\tilde{\mathcal{G}}_y = P(Y, \pi/2 + 10^{-1})$, $\tilde{\mathcal{G}}_{CZ} = P(Z^1 Z^2 - Z^1 - Z^2, \pi/2 + 10^{-1})$ (see eq. (2.30)). The different colors portray choices of basis; the yellow circles $\mathcal{V} = \mathbb{I}$, the blue stars $\mathcal{V} = \mathcal{U}$ where $\mathcal{U}$ is found through eq. (4.17), and the green squares $\mathcal{V} = \mathcal{U}^2$. Here the lines correspond to the fit for sequence lengths of m=5 to 10. The choice $\mathcal{V} = \mathcal{U}$ produces the evolution prescribed by proposition 1, which through extrapolation has an intercept of 1.

of $\boldsymbol{\mathcal{T}}$,

$$\begin{align*}
|\lambda - 0| &\leq \|\mathbb{E}[\boldsymbol{\mathcal{G}}_{\mathrm{tr}} \otimes \boldsymbol{\delta}_{\mathcal{G}}^{(I)} \boldsymbol{\mathcal{G}}]\|_2 \tag*{(Bauer-Fike)} \\
&\leq \mathbb{E}\|[\boldsymbol{\mathcal{G}}_{\mathrm{tr}} \otimes \boldsymbol{\delta}_{\mathcal{G}}^{(I)} \boldsymbol{\mathcal{G}}]\|_2 \tag*{(triangle ineq.)} \\
&= \mathbb{E}\|\boldsymbol{\delta}_{\mathcal{G}}^{(I)}\|_2 \tag*{(Unitary invariance)} \\
&\leq O\left(\mathbb{E}\sqrt{r(\mathbb{I} + \delta_{\mathcal{G}}^{(I)})}\right) \tag*{([166])} \\
&\leq O\left(\sqrt{r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})}\right) \tag*{(concavity)}
\end{align*}$$

This spectral profile implies that $(\boldsymbol{\mathcal{T}}/p)^n$ converges quickly to a rank-1 operator since the eigenvalues close to zero are exponentially suppressed.

Hence, we can approximate the asymptotic eigen-operators defined in eqs. (4.24a) and (4.24b) as:

$$\boldsymbol{\mathcal{A}}_\infty = \boldsymbol{\mathcal{A}}_4 + O(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2) \ , \tag{4.37a}$$

$$\boldsymbol{\mathcal{B}}_\infty = \boldsymbol{\mathcal{B}}_4 + O(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2) \ . \tag{4.37b}$$

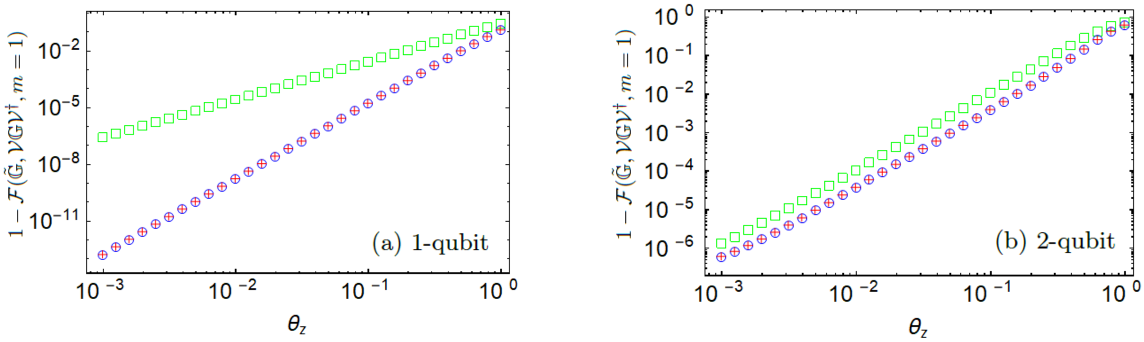


Figure 4.3: $1-\mathcal{F}(\tilde{\mathbb{G}}, \mathcal{V}\mathbb{G}\mathcal{V}^\dagger, m=1)$ as function of the angle θ_z in noise model generated by $\tilde{\mathcal{G}}_x = P(\sigma_z, \theta_z)\mathcal{G}_x$ and $\tilde{\mathcal{G}}_y = P(\sigma_z, \theta_z)\mathcal{G}_y$, $\tilde{\mathcal{G}}_{CZ} = P(\sigma_z^1\sigma_z^2 - \sigma_z^1 - \sigma_z^2, \pi/2 + 10^{-1})$ (see eq. (2.30)), with $\mathcal{V} = \mathbb{I}$ (green squares) and $\mathcal{V} = \mathcal{U}$ (blue circles) where $\mathcal{U}$ is found through eq. (4.17). The red crosses correspond to $(1-p)/2$ obtained through RB experiments.

In the simple noise model $\mathcal{E}_L\mathbb{G}\mathcal{E}_R$, $\boldsymbol{\mathcal{A}}_\infty \propto \boldsymbol{\Pi}_{\text{tr}}\boldsymbol{\mathcal{E}}_R$ and $\boldsymbol{\mathcal{B}}_\infty \propto \boldsymbol{\mathcal{E}}_L\boldsymbol{\Pi}_{\text{tr}}$. To pursue the analogy, we denote the m^{th} order right and left errors as

$$\boldsymbol{\mathcal{E}}_R^{(m)} = \mathbb{E}\left[(\boldsymbol{\mathcal{G}}_{m:1})^\dagger\tilde{\boldsymbol{\mathcal{G}}}_{m:1}\right] , \tag{4.38a}$$

$$\boldsymbol{\mathcal{E}}_L^{(m)} = \mathbb{E}\left[\tilde{\boldsymbol{\mathcal{G}}}_{m:1}(\boldsymbol{\mathcal{G}}_{m:1})^\dagger\right] . \tag{4.38b}$$

Combining eq. (4.38) and eq. (4.37), we get

$$\boldsymbol{\mathcal{A}}_\infty \propto \boldsymbol{\Pi}_{\text{tr}}\boldsymbol{\mathcal{E}}_R^{(4)} + O(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2) , \tag{4.39a}$$

$$\boldsymbol{\mathcal{B}}_\infty \propto \boldsymbol{\mathcal{E}}_L^{(4)}\boldsymbol{\Pi}_{\text{tr}} + O(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2) . \tag{4.39b}$$

The structure of single-qubit error channels allows us to pursue a deeper analysis. It follows from the channel analysis provided in [145] that, for high-fidelity qubit-channels, the 3×3 submatrix acting on the traceless hyperplane can always be decomposed as

$$\boldsymbol{\mathcal{E}}\boldsymbol{\Pi}_{\text{tr}} = \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{V}}\boldsymbol{\Pi}_{\text{tr}} \tag{4.40}$$

where $\mathcal{V}$ is a physical unitary, and $\mathcal{D}$ is an incoherent process. Here we label a channel $\mathcal{D}$ incoherent if

$$\frac{\langle \boldsymbol{\Pi}_{\text{tr}}, \boldsymbol{\mathcal{D}}\rangle}{\|\boldsymbol{\Pi}_{\text{tr}}\|_F^2} = \frac{\|\boldsymbol{\mathcal{D}}\boldsymbol{\Pi}_{\text{tr}}\|_F}{\|\boldsymbol{\Pi}_{\text{tr}}\|_F} + O(r(\mathcal{D})^2) . \tag{4.41}$$

Incoherent channels have the additional property that , given an error channel $\boldsymbol{\Lambda}$ [33]

$$\frac{\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{D}}\boldsymbol{\Lambda}\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} = \frac{\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{D}}\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2}\frac{\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\Lambda}\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F^2} + O(r(\mathcal{D}\Lambda)^2)\ . \tag{4.42}$$

Expressing the 4th order right error $\boldsymbol{\mathcal{E}}_R^{(4)}$ as

$$\boldsymbol{\mathcal{E}}_R^{(4)}\boldsymbol{\Pi}_{\mathrm{tr}} = \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{V}}\boldsymbol{\Pi}_{\mathrm{tr}}\ . \tag{4.43}$$

allows us to maximally correct it through a physical unitary:

$$F(\mathcal{E}_R^{(4)}\mathcal{V}^\dagger, \mathbb{I}) = \max_{\mathcal{U}} F(\mathcal{E}_R^{(4)}\mathcal{U}, \mathbb{I}) \geq F(\mathcal{E}_R^{(4)}, \mathbb{I})\ . \tag{4.44}$$

Using the property expressed in eq. (4.42), we get:

$$\begin{aligned}\frac{\left\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{E}}_R^{(4)}\boldsymbol{\mathcal{V}}^\dagger\boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{E}}_L^{(4)}\right\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F} =& \frac{\left\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{E}}_R^{(4)}\boldsymbol{\mathcal{V}}^\dagger\right\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F}\frac{\left\langle \boldsymbol{\Pi}_{\mathrm{tr}}, \boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{E}}_L^{(4)}\right\rangle}{\|\boldsymbol{\Pi}_{\mathrm{tr}}\|_F} \\ &+ O(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2).\end{aligned} \tag{4.45}$$

Looking back at theorem 25 and using eqs. (4.39a), (4.39b) and (4.45) results in

$$C(\mathcal{V}^\dagger) = 1 + O\left(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2\right)\ . \tag{4.46}$$

Since both $\mathcal{V}$ and $\mathcal{E}_L^{(4)}$ have at most infidelity of order $r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})$, it follows that the composition $\mathcal{V}\mathcal{E}_L^{(4)}$ must also have an infidelity of order $r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})$, which guarantees

$$\sqrt{1 - b^2(\mathcal{V}^\dagger)} = O\left(\sqrt{r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})}\right), \tag{4.47}$$

while incoherence guarantees

$$\sqrt{1 - a^2(\mathcal{V}^\dagger)} = O\left(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})\right). \tag{4.48}$$

Using

$$|w(\mathcal{V}^\dagger)^\dagger \boldsymbol{\Delta} v(\mathcal{V}^\dagger)| \leq \mathbb{E}\|\delta_{\mathcal{G}}^{(I)}\|_2 \leq O\left(\sqrt{r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})}\right) \tag{4.49}$$

in eq. (4.33b), we find

$$D(1, \mathcal{V}^\dagger) = O\left(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2\right) , \tag{4.50}$$

which, together with eqs. (4.32) and (4.46) leads to

$$f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{V}^\dagger \mathbb{G} \mathcal{V}, m) = p^m + O\left(r(\mathbb{I} + \mathbb{E}\delta_{\mathcal{G}}^{(I)})^2\right) . \tag{4.51}$$

This expression allows us to pick a better perturbative expansion than eq. (4.35). Indeed, choosing

$$\boldsymbol{\tilde{\mathcal{G}}} = \boldsymbol{\mathcal{V}^\dagger \mathcal{G} \mathcal{V}} + \boldsymbol{\delta_{\mathcal{G}}^{(V^\dagger)}}\ \boldsymbol{\mathcal{V}^\dagger \mathcal{G} \mathcal{V}} , \tag{4.52}$$

ensures that the noisy operations $\mathbb{I} + \delta_{\mathcal{G}}^{(V^\dagger)}$ have an gate-set circuit infidelity which is more in line with the RB data:

$$r(\mathbb{I} + \delta_{\mathcal{G}}^{(V^\dagger)}) = \frac{d-1}{d}(1-p) + O(r(\mathbb{I} + \delta_{\mathcal{G}}^{(I)})^2) . \tag{4.53}$$

Iterating the analysis leads to

$$f_{\mathrm{tr}}(\tilde{\mathbb{G}}, \mathcal{V}^\dagger \mathbb{G} \mathcal{V}, m) = p^m + O\left((1-p)^2\right) . \tag{4.54}$$

This completes the demonstration of proposition 1.

Our current proof technique relies on the structure of single-qubit channels. For higher dimensions, we conjecture that an analog of proposition 1 holds, although the scaling with the dimension is unclear.

Conjecture 26. If the fidelity of $\mathcal{E}_R^{(4)}$ is high, then $\exists$ a physical unitary $\mathcal{V}^\dagger$ s.t. $\mathcal{E}_R^{(4)}\mathcal{V}^\dagger$ is incoherent.

As we now show constructively, conjecture 26 holds for physically motivated noise models composed of generalized dephasing, amplitude damping, and unitary processes. Under such noise models,

$$\mathcal{E}_R^{(4)} = \mathcal{U}_T \mathcal{D}_T \cdots \mathcal{U}_2 \mathcal{D}_2 \mathcal{U}_1 \mathcal{D}_1 \tag{4.55}$$

for some unitaries $\mathcal{U}_i$ and incoherent channels $\mathcal{D}_i$.

The channel $\mathcal{U}\mathcal{D}\mathcal{U}^\dagger$ is incoherent for any physical unitary $\mathcal{U}$, and the composition of incoherent channels is also incoherent, so eq. (4.55) can be rewritten as $\mathcal{E}_R^{(4)} = \mathcal{D}\mathcal{V}$, where $\mathcal{D}$ and $\mathcal{V}$ are incoherent and unitary respectively:

$$\mathcal{D} = (\mathcal{U}_T \mathcal{D}_T {\mathcal{U}_T}^\dagger) \cdots (\mathcal{U}_{T:1} \mathcal{D}_1 {\mathcal{U}_{T:1}}^\dagger) \tag{4.56}$$

$$\mathcal{V} = \mathcal{U}_{T:1} . \tag{4.57}$$

Acknowledgments

The authors acknowledge helpful discussions with Timothy J. Proctor and Hui Khoon Ng. This research was supported by the U.S. Army Research Office through grant W911NF-14-1-0103. This research was undertaken thanks in part to funding from TQT, CIFAR, the Government of Ontario, and the Government of Canada through CFREF, NSERC and Industry Canada.

Chapter 5

Randomized benchmarking under different gate sets

5.1 Overview

We provide a comprehensive analysis of the differences between two important standards for randomized benchmarking (RB): the Clifford-group RB protocol proposed originally in Emerson et al (2005) and Dankert et al (2006), and a variant of that RB protocol proposed later by the NIST group in Knill et al, PRA (2008). While these two protocols are frequently conflated or presumed equivalent, we prove that they produce distinct exponential fidelity decays leading to differences of up to a factor of 3 in the estimated error rates under experimentally realistic conditions. These differences arise because the NIST RB protocol does not satisfy the unitary two-design condition for the twirl in the Clifford-group protocol and thus the decay rate depends on non-invariant features of the error model. Our analysis provides an important first step towards developing definitive standards for benchmarking quantum gates and a more rigorous theoretical underpinning for the NIST protocol and other RB protocols lacking a group-structure. We conclude by discussing the potential impact of these differences for estimating fault-tolerant overheads.

5.2 Introduction

Clifford-group randomized benchmarking (RB) [49, 45] has become the *de facto* standard tool for assessing and optimizing the quantum control required for quantum computing systems by estimating error rates associated with sets of elementary gates operations. It has been known for some time that this protocol leads to an invariant exponential decay [49, 45, 99] because it is equivalent to a sequence of twirls [99] with unitary-two designs [45].

More recently, the robustness of the Clifford-group RB protocol has been supported by a rigorous theoretical framework, including proofs that an exponential fidelity decay will be observed under very broad experimental conditions. This includes essentially arbitrary state preparation and measurement errors [106, 107] and gate-dependent errors [171], as well as proofs that the observed error rate relates directly to a well-defined notion of gate-fidelity [171, 131, 31], which has fully overcome recent concerns about relating measured RB error rates to a meaningful concept of gate-fidelity under gate-dependent errors [131].

While a wide-variety of group-based generalizations of RB have been proposed in recent years, *e.g.* [12, 32, 43, 173, 172, 171], in this chapter we focus on clarifying the physical relevance of a standing conflation in the literature between the now standard Clifford-group RB protocol proposed in [49, 45] and an alternate version of RB proposed later by NIST [92]. As described below, these are distinct protocols that measure distinct properties of the error model and thus can produce different error rate estimates under the same, realistic experimental conditions. Moreover, because the NIST protocol does not admit a closed-group or unitary two-design structure, the rigorous theoretical framework justifying Clifford-group RB does not trivially extend to support the physical interpretation and robustness of NIST RB.

In this Letter we identify the operationally-relevant differences between the Clifford-group RB protocol and the NIST version of RB which clarifies how they can lead to very different error rate estimates given the same error model (as defined in terms of the elementary control pulses). We then provide the first rigorous proof that the NIST RB protocol does indeed produce an exponential decay under gate-independent error models. This is an important step toward developing a theoretical justification for the NIST protocol and other RB protocols that do not admit a group-structure in the case of gate-dependent errors and the ultimate goal of a theoretical framework within which error reconstruction under RB protocols with different gate sets can be extracted in a unified and consistent manner. Our analysis is thus also essential for comparing cross-platform benchmarking methods and standards for quantum computing.

As a second contribution, we numerically explore the size and scope of the quantitive differences in estimated error rates that arise under each of the protocols for a variety of physically relevant error models and pulse-decompositions, and observe that experimentally estimated error rates can differ by as much as a factor of 3 in typical cases. We conclude by discussing how these differences are relevant for detecting gate-dependent errors, and estimating fault-tolerant overheads under quantum error correction.

5.3 Background and Motivation

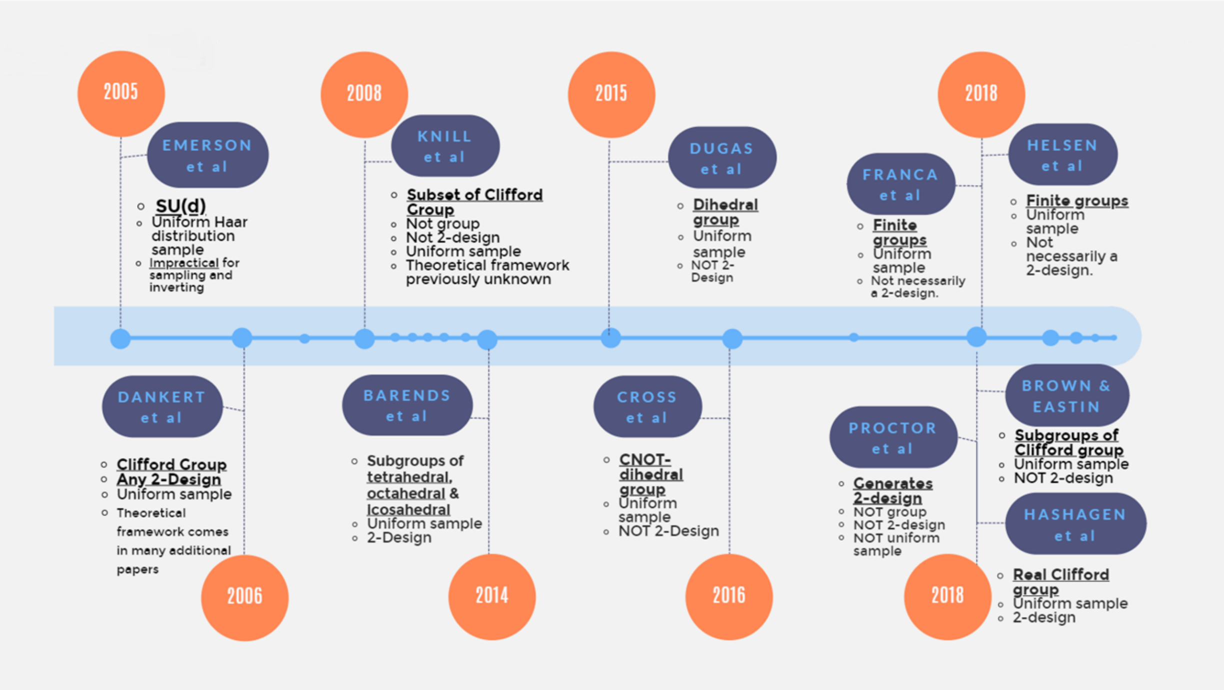


Figure 5.1: This timeline gives a clear overview of papers exploring the advantages and disadvantages of Randomized Benchmarking under different gate sets. From its development in 2005 with the use of SU(d)[49], to the more recent works which utilize other gate sets to probe various aspects of the error model[45, 92, 10, 32, 43, 77, 60, 27, 133, 74].

The original proposal for randomized benchmarking from Emerson *et al.* [49] considered implementing long sequences of quantum gates drawn *uniformly* at random from the group SU(d) to characterize quantum systems with Hilbert space dimension d.

Protocol 5: Standard Clifford-group RB, as described in [49, 45].

1. Sample a set of m gates $\mathcal{G}_i$ picked independently and uniformly at random from the Clifford group $\mathbb{C}$ defined in eq. (5.5);
2. Determine the recovery gate $\mathcal{G}_{m+1}$ (see text below);
3. Prepare a state $\rho \approx |0\rangle\langle 0|$;
4. Perform the sampled gates from step 1, followed by the recovery gate $\mathcal{G}_{m+1}$ determined in step 2: $\tilde{\mathcal{G}}_{m+1:1} = \tilde{\mathcal{G}}_{m+1} \circ \ldots \circ \tilde{\mathcal{G}}_1$;
5. Measure a POVM $\{Q, \mathbb{I} - Q\}$, where the first observable is $Q \approx \mathcal{G}_{m+1:1}(|0\rangle\langle 0|)$, and respective outcome labels are {"recovery", "non-recovery"};
6. Repeat steps 3–5 a number times to estimate the probability of observing the "recovery" event $\Pr(\text{“recovery”}|\{\mathcal{G}_i\}, m) = \operatorname{tr} Q\tilde{\mathcal{G}}_{m+1:1}(\rho)$;
7. Repeat steps 1–6 for s different sets of m randomly sampled gates $\{\mathcal{G}_i\}$;
8. Repeat steps for 1–7 for different values of m of random gates.
9. Fit the estimated recovery probabilities to the decay model
$$A_{\mathbb{C}}\, p_{\mathbb{C}}^m + B_{\mathbb{C}} \; ; \tag{5.1}$$
10. Estimate the Clifford gate-set infidelity through
$$r_{\mathbb{C}} = (1 - p_{\mathbb{C}})/2 \; . \tag{5.2}$$

That work proved that the measured fidelity would follow an exponential decay with a decay rate that is fixed uniquely by the error model, that is, the measured decay rate would not depend on the choice of initial state or the specific random quantum gate sequences.

This protocol suffered from two limitations: the random gates were drawn from a continuous set, which is impractical even for $d = 2$, and the protocol would not be efficient for large systems because a typical random element of SU(d) requires exponentially long gate sequences under increasing numbers of qubits. Additionally, in that limit the inversion gate may not be computed efficiently.

However, practical and efficient solutions to both of these problems were proposed in Dankert *et al.* [45] in 2006, which proved and observed that drawing gates uniformly at random from the Clifford group would lead to the same exponential decay rate as computed in the protocol proposed earlier in Emerson *et al.* [49], which follows from the unitary 2-design property of the Clifford group.

This connection is made more explicit through the observation that a random sequence of gates drawn from any group is equivalent to an independent sequence of twirls under that group, as shown explicitly in [99] and had been conjectured earlier in [49].

Collectively these papers define what is now known as Clifford-group RB, an efficient and practical method for assessing error rates for quantum processors on arbitrarily large numbers of qubits, summarized here as Protocol 5. This Clifford-group RB protocol has become a *de facto* standard for benchmarking and optimizing gate performance and has been implemented by a large number of groups across various hardware platforms to characterize single- and multi-qubit gate operations, see, *e.g.*, Refs [181, 119, 86, 10, 113, 152, 157].

The theoretical underpinnings of the standard protocol were clarified and further developed by Magesan *et al.* [106, 107], which showed that the exponential decay rate was robust to state preparation and measurement errors (SPAM), and by Wallman [171] and Dugas *et al.* [31], which showed that the exponential decay rate was meaningfully related to a gate-fidelity in spite of the gauge freedom highlighted by Proctor *et al.* [131] that occurs in the usual definition of the average gate-fidelity.

Additionally, the work of Wallman [171] established that the RB error rate is robust to very large variations in the error model over the gate set (known as gate-dependent error models) and thus established that RB can also be an effective tool for diagnosing non-Markovian errors. This follows from the fact that only non-Markovian errors (including what are sometimes called time-dependent Markovian errors) can produce a statistically significant deviation from an exponential decay under a Clifford-group RB experiment.

A different version of the 2005 Emerson *et al.* [49] protocol was proposed by Knill *et al.* [92] in 2008 and implemented in the NIST ion trap. This proposal involved the same kind of motion reversal experiment proposed in Emerson et al [49] but selects random sequences of gates drawn from a *non-uniform* sampling of the single-qubit Cliffords, defined as "Pauli-randomized $\pi/2$ gates". The precise recipe for this protocol is summarized as Protocol 6. The NIST version of the randomized benchmarking protocol continues to be implemented mainly in ion traps [26, 73]. We note that in contrast to the earlier Clifford-group RB protocol which is defined for single- and multi-qubit gate operations, the NIST version of RB is defined only for single-qubit gate operations.

Protocol 6: NIST RB, as described in [92].

---

1. Sample a set of m gates $\mathcal{G}_i$ picked independently and uniformly at random from the NIST gate-set $\mathbb{N}$ defined in eq. (5.8);

2–8. Idem as in protocol 5.

9. Fit the estimated recovery probabilities to the decay model

$$A_{\mathbb{N}}\, p_{\mathbb{N}}^{m} + B_{\mathbb{N}} \ . \tag{5.3}$$

10. Estimate the NIST gate-set infidelity through

$$r_{\mathbb{N}} = (1 - p_{\mathbb{N}})/2 \ . \tag{5.4}$$

In this Letter, we prove that the measured fidelity under the NIST protocol will follow an exponential decay, which has never been established for this protocol, and relate the decay rate to the intrinsic properties of the error model, demonstrating how it differs from the properties measured by Clifford-group RB. This analysis also provides the first step towards developing a self-consistent theoretical framework for interpreting and relating the results of the large and growing family of RB-style protocols, which all share the structure of applying random sequences of gates and differ mainly through the choice of random gate-set [32, 43, 173, 180, 37, 172, 64, 12, 133, 89, 50, 107, 152, 171].

Finally, an additional motivation for the present work comes from the recent conceptual development [174] establishing how accurately RB error estimation methods can inform the design and 'in situ' performance of large-scale quantum computations. This development overcomes a standing criticism of RB protocols that the very nature of a randomization protocol limits these protocols to detect only the stochastic component of coherent errors - and hence that RB-type protocols are not able to capture the full impact of these errors. Coherent errors are those that typically arise from imperfect quantum control due to residual mis-calibrations [1] and pose a major challenge for reliable quantum computation. However, this perceived limitation has become a strength of RB protocols thanks to the concept of randomized compiling [174].

Randomized compiling is an important generalization and improvement to the concept of Pauli-Frame Randomization (PFR) proposed earlier in [94] that does not require any

---

[1] Note that cross-talk is a non-trivial coherent error that results from control errors affecting distant qubits.

overhead for the randomization and works for universal gate sets [2]. When implementing a quantum algorithm via randomized compiling, the only performance limiting component of a coherent error is precisely the stochastic component that is detected via RB protocols. In summary, a precise and accurate understanding of RB error estimates is highly relevant because RB detects precisely the component of the error that determines the 'in vivo' performance of the gate operations within a large-scale circuit performed via randomized compiling.

5.4 Results

5.4.1 Standard RB vs NIST RB

The standard RB protocol (SRB) [49, 45] is summarized in protocol 5. The recovery operations mentioned in step 2 is usually an inversion gate, $\mathcal{G}_{m+1} = \mathcal{G}_{m:1}^{-1}$, in which case the recovery observable simply corresponds to the initial state: $Q \approx |0\rangle\langle 0|$. However, performing the inverse only up to a random bit flip (i.e $\mathcal{G}_{m+1} = X_\pi^b \mathcal{G}_{m:1}^{-1}$) leads to a simpler decay model with fewer free parameters because then $B = 1/2$. Of course in this case one has to keep track of the bit flip, that is $Q \approx X_\pi^b(|0\rangle\langle 0|)$. Such a randomized recovery operation was proposed in [92].

SRB is typically implemented using the Clifford group $\mathbb{C}$ as a randomizing gate-set, as specified in the first step of protocol 5, but the derivation of the decay model shown in eq. (5.1) holds for any unitary 2-design [45]. The Clifford group is defined as follows. First consider pulses along any Cartesian axis system

$$X_\theta := e^{-i\theta/2\ X},\ Y_\theta := e^{-i\theta/2\ Y},\ Z_\theta := e^{-i\theta/2\ Z},$$

where σ_i denote the unitary Pauli matrices. The Pauli group $\mathbb{P}$ is defined in terms of the identity operation and 3 elementary π pulses: $\mathbb{P} := \{\mathbb{I}, X_\pi, Y_\pi, Z_\pi\}$.

The Clifford group $\mathbb{C}$ is the normalizer of the Pauli group and can be obtained from the Pauli group composed with the coset $\mathbb{S} := \{\mathbb{I}, X_{\pi/2}, Y_{\pi/2}, Z_{\pi/2}, Z_{\pi/2}X_{\pi/2}, X_{-\pi/2}Z_{-\pi/2}\}$:

$$\mathbb{C} := \mathbb{S} \cdot \mathbb{P} = \{\mathcal{S} \circ \mathcal{P} \mid \mathcal{S} \in \mathbb{S}, \mathcal{P} \in \mathbb{P}\}\ . \tag{5.5}$$

---

[2] In particular, relative to PFR, randomized compiling (i) does not add additional overhead to each clock cycle, which it achieves by 'compiling in' the randomizing gates, (ii) works for universal gate sets, and (iii) rigorously characterizes how close the effective error model is to a purely stochastic error model under errors gate-dependent errors

Some other experimental groups performed RB using alternate 2-design gate-sets in step 1 of protocol 5[10]. Amongst the set of possible unitary 2-designs, it is worth mentioning those following subsets of $\mathbb{C}$. Consider the cyclic group $\mathbb{T} := \{\mathbb{I}, Z_{\pi/2}X_{\pi/2}, X_{-\pi/2}Z_{-\pi/2}\}$, then the following sets both form 2-designs of order 12:

$$\mathbb{C}_{12} := \mathbb{T} \cdot \mathbb{P} = \{\mathcal{T} \circ \mathcal{P} \mid \mathcal{T} \in \mathbb{T}, \mathcal{P} \in \mathbb{P}\} \ , \tag{5.6}$$

$$\sqrt{Z}\mathbb{C}_{12} := Z_{\pi/2} \cdot \mathbb{C}_{12} = \{Z_{\pi/2} \circ \mathcal{C} \mid \mathcal{C} \in \mathbb{C}_{12}\} \ , \tag{5.7}$$

with $\mathbb{C}_{12} \cup \sqrt{Z}\mathbb{C}_{12} = \mathbb{C}$. Obviously, the decay parameters as well as the infidelity depend on the randomizing gate-set (hence the indices).

The validity of the decay model and the connection between the decay parameter and the gate-set infidelity have been demonstrated in the case of gate-independent Markovian noise scenarios in [49]. The proofs of eq. (5.1) and eq. (5.2) have been generalized to encompass gate-dependent noise scenarios in [171, 116] and [31] respectively[3].

Although the proof techniques can get mathematically heavy, their essence remains simple: the algebraic richness of 2-designs prevents the unpredictable accumulation of errors with increasing circuit length. As we show with more care in the next section, random sampling over the gate-set tailors the effective errors at each cycle to a depolarizing channel for which the evolution is parameterized by a single real number p. The errors are stripped of all properties except one, which turns out to be in one-to-one correspondence with their average infidelity. By modifying the sampled circuits lengths, we can estimate the parameter p and retrieve the infidelity.

While unitary 2-designs are provably effective randomizing gate-sets, leading to the model portrayed in eq. (5.1), some algebraically weaker gate-sets have indicated a similar exponential decaying behaviour.

The gate-set $\mathbb{N}$ used in NIST RB [92] is a composition of a set $\mathbb{Q} := \{X_{\pm\pi/2}, Y_{\pm\pi/2}\}$, consisting of $\pi/2$ pulses in the xy-plane, with the Pauli operators:

$$\mathbb{N} := \mathbb{Q} \cdot \mathbb{P} = \{\mathcal{Q} \circ \mathcal{P} \mid \mathcal{Q} \in \mathbb{Q}, \mathcal{P} \in \mathbb{P}\} \ . \tag{5.8}$$

$\mathbb{N}$ has order 8, and although it contains all its inverse elements (that is $\forall \mathcal{N} \in \mathbb{N}$, $\exists \mathcal{M} \in \mathbb{N}$ s.t. $\mathcal{M} \cdot \mathcal{N} = \mathbb{I}$), it is not closed under multiplication. It does not form a group, nor a 2-design; however, the closure $\langle \mathbb{N} \rangle$ forms the Clifford group $\mathbb{C}$.

---

[3] In gate-dependent noise scenarios, the connection between the RB decay parameter and the gate-set infidelity remains a (strongly supported) conjecture for $d > 2$.

RB sequences can be seen as Markov chains [114], where the elements of the chain are the aggregate circuits, that is $\mathcal{C}_1 = \mathcal{G}_1$, $\mathcal{C}_2 = \mathcal{G}_2\mathcal{G}_1$, $\mathcal{C}_m = \mathcal{G}_{m:1}$. Indeed, the probability distribution on circuits $\mathcal{C}_m = \mathcal{G}_{m:1}$ simply depends on the circuit $\mathcal{C}_{m-1}$ and on the probability distribution of the random gate applied at step m. In standard RB, $\mathcal{C}_i$ is always uniformly distributed over the Clifford group. In NIST RB, $\mathcal{C}_{2n}$ (or $\mathcal{C}_{2n+1}$) converges to a uniform distribution over $\mathbb{C}_{12}$ (or $\sqrt{Z}\mathbb{C}_{12}$), as shown in fig. 5.2.

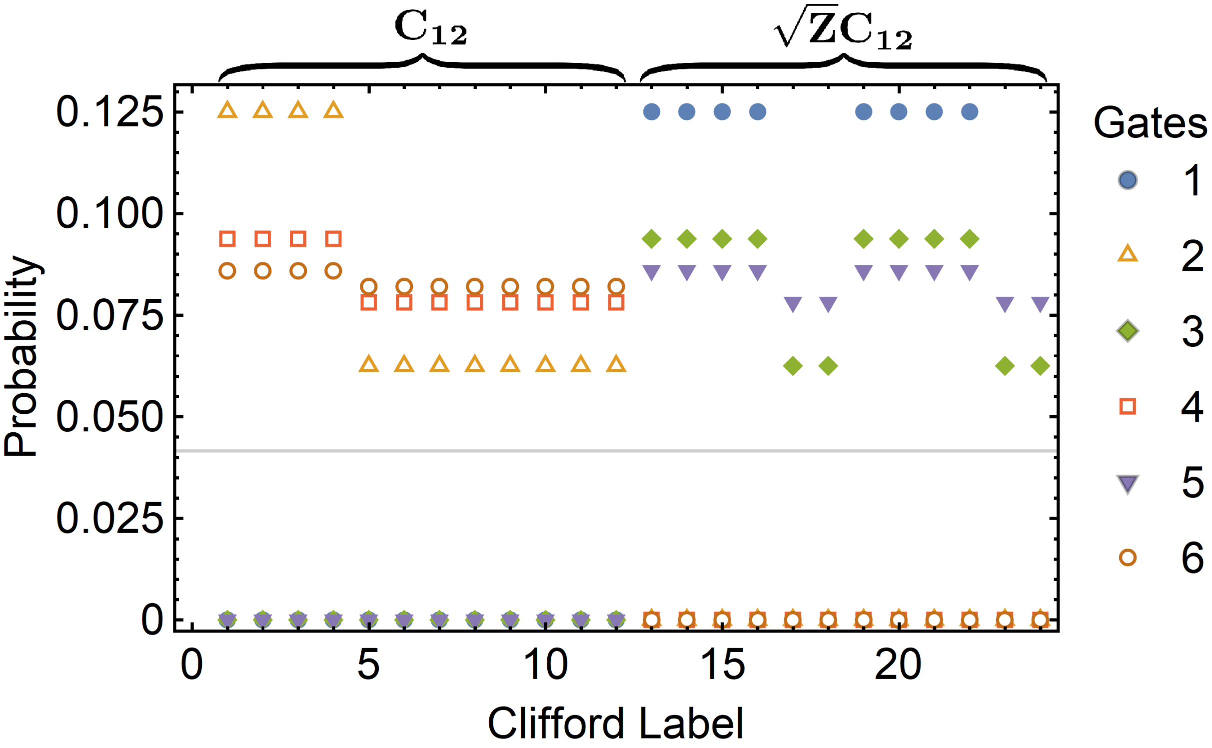


Figure 5.2: Probability distribution over the Clifford gates $\mathbb{C}$ (labelled as in [10]) after m gates (i.e clock cycles) of NIST RB drawn uniformly at random from $\mathbb{N} \subset \mathbb{C}$. This leads to a *non-uniform* sampling over the Cliffords that varies as m increases. Asymptotically, for a sequence of even (or odd) length (hollow (or solid) shapes), the probability distribution tends toward a uniform distribution over $\mathbb{C}_{12}$ (or $\sqrt{Z}\mathbb{C}_{12}$). The grey line indicates an equal probability over the full 24 Clifford group $\mathbb{C}$.

While this approach to RB has been useful for estimating error rates [92, 26, 73], in the absence of a unitary 2-design structure, it is not clear how to relate the measured probabilities from protocol 6 to the usual decay predicted under SRB, or to any definition of infidelity for that matter. In this paper we provide a concrete analysis of the outcome of protocol 6, which yields a justification and interpretation for the decay model eqs. (5.3) and (5.4).

It is important to emphasize that NIST RB now falls into a family of RB protocols defined as "direct RB" [133]. The analysis below gives a concrete instance of direct RB that both justifies and interprets past experiments and gives an insightful example of the main idea behind direct RB.

5.4.2 Theoretical Analysis of NIST RB

The goal of this section is to provide the key insight behind the mechanics of NIST RB. To lighten up the mathematical machinery, we assume a gate-independent error model[4], where the noisy gates are followed by an error $\mathcal{E}$:

$$\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G} \ . \tag{5.9}$$

In such a model, the gate-set infidelities $r_{\mathbb{C}}$ and $r_{\mathbb{N}}$ are *de facto* equal to the infidelity of the error $r(\mathcal{E}, \mathbb{I})$. We show that $r(\mathcal{E}, \mathbb{I})$ can be estimated by both protocols 5 and 6.

The recovery probabilities look like

$$\mathrm{Tr}\, \mathcal{Q}\mathcal{E}\, \underbrace{X^b_\pi \mathcal{G}^{-1}_{m:1}}_{\mathcal{G}_{m+1}} \mathcal{E}\mathcal{G}_m \cdots \mathcal{E}\mathcal{G}_2\mathcal{E}\mathcal{G}_1(\rho) \ . \tag{5.10}$$

Compiling the last error $\mathcal{E}$ as well as the random bit flip X^b_π into the measurement procedure (that is, letting $Q \to X^b_\pi \mathcal{E}^\dagger(Q)$), leaves us with the random sequence which is at the heart of both NIST RB and SRB protocols:

$$S(\{\mathcal{G}_i\}) = \mathcal{G}^{-1}_{m:1}\mathcal{E}\mathcal{G}_m \cdots \mathcal{E}\mathcal{G}_2\mathcal{E}\mathcal{G}_1 \ . \tag{5.11}$$

In SRB, the next step in the analysis is to redefine the gates as $\mathcal{G}_i = \mathcal{G}'_i\mathcal{G}'^{-1}_{i-1}$ (with $\mathcal{G}_1 = \mathcal{G}'_1$), where both $\mathcal{G}_i$ and $\mathcal{G}'_i$ are picked uniformly at random from the randomizing set. Such a relabeling is possible because the randomizing gate-set is usually a group. Averaging over all sequences yields

$$\begin{aligned} \mathbb{E}_{\{\mathcal{G}_i\}} S(\{\mathcal{G}_i\}) &= \mathbb{E}_{\{G'_i\}} \mathcal{G}'^{-1}_m \mathcal{E}\mathcal{G}'_m \cdots \mathcal{G}'^{-1}_2 \mathcal{E}\mathcal{G}'_2 \mathcal{G}'^{-1}_1 \mathcal{E}\mathcal{G}'_1 \\ &= \left(\mathcal{E}^{\mathbb{C}}\right)^m \ , \end{aligned} \tag{5.12}$$

---

[4] The formal analysis of direct RB under more general gate-dependent noise scenarios will be considered in subsequent work. However, in section 5.6.1 we elucidate the broad reasoning required for a gate-dependent analysis.

where

$$\mathcal{E}^{\mathbb{C}} := \frac{1}{|\mathbb{C}|} \sum_{\mathcal{G} \in \mathbb{C}} \mathcal{G}^{-1} \mathcal{E} \mathcal{G} \tag{5.13}$$

is referred to as the twirl of the error $\mathcal{E}$ over the gate-set $\mathbb{C}$. If $\mathbb{C}$ is a 2-design, then the twirled channel $\mathcal{E}^{\mathbb{C}}$ is reduced to a depolarizing channel. To mathematically concretize the description of a channel $\mathcal{E}$, we resort to the 4×4 Pauli-Liouville representation, which is defined as

$$\mathcal{E}_{ij} := \frac{1}{2} \operatorname{tr} B_j^\dagger \mathcal{E}(B_i) \tag{5.14}$$

where $B_1 = \mathbb{I}$, $B_2 = X$, $B_3 = Y$, $B_4 = Z$. In this representation, the depolarizing channel $\mathcal{E}^{\mathbb{C}}$ is expressed as a diagonal matrix $\operatorname{diag}(1, p_{\mathbb{C}}, p_{\mathbb{C}}, p_{\mathbb{C}})$, where $p_{\mathbb{C}}$ is a real number close to 1:

$$p_{\mathbb{C}} = \frac{\mathcal{E}_{22} + \mathcal{E}_{33} + \mathcal{E}_{44}}{3} \ . \tag{5.15}$$

The averaged core sequence hence evolves as

$$\mathbb{E}_{\{\mathcal{G}_i\}} S(\{\mathcal{G}_i\}) = \left(\mathcal{E}^{\mathbb{C}}\right)^m = \operatorname{diag}(1, p_{\mathbb{C}}^m, p_{\mathbb{C}}^m, p_{\mathbb{C}}^m) \ . \tag{5.16}$$

Deriving eq. (5.1) is then simply a matter of incorporating SPAM procedures in the evaluation of the recovery probabilities. Straightforward algebra links the infidelity of $\mathcal{E}$ with its diagonal Liouville matrix elements through

$$r(\mathcal{E}, \mathbb{I}) = \frac{1}{2} - \frac{\mathcal{E}_{22} + \mathcal{E}_{33} + \mathcal{E}_{44}}{6} \ . \tag{5.17}$$

The relation between the decay constant $p_{\mathbb{C}}$ and the gate-set infidelity $r_{\mathbb{C}} = r(\mathcal{E}, \mathbb{I})$ results from combining eq. (5.15) and eq. (5.17).

The relabeling trick resulting in a m-composite depolarizing channel is not possible in NIST RB: $\mathbb{N}$ is neither a group nor a 2-design. However, although $\mathbb{N}$ has a weaker algebraic structure, it is not completely devoid of interesting properties. Indeed, every element of $\mathbb{N}$ can be written as $\mathcal{P}_{\text{left}} \cdot \mathcal{Q} \cdot \mathcal{P}_{\text{right}}$, where $\mathcal{P}_{\text{left}}$, $\mathcal{P}_{\text{right}} \in \mathbb{P}$ and $\mathcal{Q} \in \mathbb{Q} := \{X_{\pm\pi/2}, Y_{\pm\pi/2}\}$. Using this, we can relabel every gate $\mathcal{G}_i$ as

$$\mathcal{G}_1 = \mathcal{C}_1 \mathcal{Q}_1 \ , \tag{5.18a}$$

$$\mathcal{G}_i = \mathcal{C}_i \mathcal{Q}_i \mathcal{C}_{i-1}^{-1} \quad (i = 2, \cdots, m) \ , \tag{5.18b}$$

$$\mathcal{G}_{m:1}^{-1} = \mathcal{Q}_{m:1}^{-1} \mathcal{C}_m^{-1} \tag{5.18c}$$

where $\mathcal{C}_i$ is chosen UAR from the Pauli group $\mathbb{P}$, and $\mathcal{Q}_i$ are chosen UAR from $\mathbb{Q}$. Using such a manipulation and randomizing over the Paulis transform the core sequence into

$$\mathbb{E}_{\{\mathcal{C}_i\}} S(\{\mathcal{G}_i\}) = \mathcal{Q}_{m:1}^{-1}\mathcal{E}^{\mathbb{P}}\mathcal{Q}_m \cdots \mathcal{E}^{\mathbb{P}}\mathcal{Q}_2\mathcal{E}^{\mathbb{P}}\mathcal{Q}_1 \;, \tag{5.19}$$

where $\mathcal{E}^{\mathbb{P}}$ is the error channel twirled over the Pauli group. In the Pauli-Liouville picture, the Pauli group has 4 inequivalent irreps; the twirled channel is diagonal:

$$\mathcal{E}^{\mathbb{P}} = \mathrm{diag}(1, x, y, z) \;, \tag{5.20}$$

where $x = \mathcal{E}_{22}$, $y = \mathcal{E}_{33}$, $z = \mathcal{E}_{44}$. The relabeling method still can't be used with the $\mathcal{Q}_i$'s, but the simplification of the noise channel $\mathcal{E}$ through the Pauli twirl unveils a recursive approach. Consider the $m = 1$ case:

$$\mathbb{E}_{\{\mathcal{G}_1\}} S(\{\mathcal{G}_1\}) = \mathbb{E}_{\{\mathcal{Q}_1\}} \mathcal{Q}_1^{-1}\mathcal{E}^{\mathbb{P}}\mathcal{Q}_1 = \mathcal{E}^{\mathbb{N}} \;, \tag{5.21}$$

where the twirl over the NIST gate-set results in

$$\mathcal{E}^{\mathbb{N}} = \mathrm{diag}\left(1, \frac{x+z}{2}, \frac{y+z}{2}, \frac{x+y}{2}\right) \;. \tag{5.22}$$

The $m = 2$ case suggests a recursion relation:

$$\mathbb{E}_{\{\mathcal{G}_i\}} S(\{\mathcal{G}_i\}) = \left(\mathcal{E}^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\right)^{\mathbb{N}} \;, \tag{5.23a}$$

$$\left(\mathcal{E}^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\right)^{\mathbb{N}} = \mathrm{diag}(1, x_2, y_2, z_2) \;, \tag{5.23b}$$

where

$$x_2 = \frac{x^{\frac{(x+z)}{2}} + z^{\frac{(x+y)}{2}}}{2} \;, \tag{5.24a}$$

$$y_2 = \frac{y^{\frac{(y+z)}{2}} + z^{\frac{(x+y)}{2}}}{2} \;, \tag{5.24b}$$

$$z_2 = \frac{x^{\frac{(x+z)}{2}} + y^{\frac{(y+z)}{2}}}{2} \;. \tag{5.24c}$$

Indeed, the general case can be expressed as

$$\mathbb{E}_{\{\mathcal{G}_i\}} S(\{\mathcal{G}_i\}) = \left(\left(\left(\mathcal{E}^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\right)^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\right)^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\cdots\right)^{\mathbb{N}} \;, \tag{5.25a}$$

$$\left(\left(\left(\mathcal{E}^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\right)^{\mathbf{N}}\mathcal{E}^{\mathbb{P}}\right)^{\mathbb{N}}\mathcal{E}^{\mathbb{P}}\cdots\right)^{\mathbb{N}} = \mathrm{diag}(1, x_m, y_m, z_m) \;, \tag{5.25b}$$

where the recursion relation can be stated as

$$x_m = \frac{x \cdot x_{m-1} + z \cdot z_{m-1}}{2} \ , \tag{5.26a}$$

$$y_m = \frac{y \cdot y_{m-1} + z \cdot z_{m-1}}{2} \ , \tag{5.26b}$$

$$z_m = \frac{x \cdot x_{m-1} + y \cdot y_{m-1}}{2} \ . \tag{5.26c}$$

Using basic linear algebra, this system of recursive equations can be expressed as

$$\begin{bmatrix} x_m \\ y_m \\ z_m \end{bmatrix} = M \begin{bmatrix} x_{m-1} \\ y_{m-1} \\ z_{m-1} \end{bmatrix} = M^m \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \ , \tag{5.27}$$

where

$$M = \frac{1}{2} \begin{bmatrix} x & 0 & z \\ 0 & y & z \\ x & y & 0 \end{bmatrix} \ . \tag{5.28}$$

x, y, and z differ from 1 by at most order $r(\mathcal{E}, \mathbb{I})$. Hence, up to the second order in the infidelity, M has the following spectrum:

$$\mathcal{E}_1 \approx \frac{x+y+z}{3} = p_{\mathbb{C}} \ , \tag{5.29a}$$

$$\mathcal{E}_2 \approx \frac{x+y}{4} \ , \tag{5.29b}$$

$$\mathcal{E}_3 \approx -\frac{x+y+4z}{12} \ . \tag{5.29c}$$

Since $\mathcal{E}_1 \approx 1$, $\mathcal{E}_2 \approx 1/2$ and $\mathcal{E}_3 \approx -1/2$, M^m converges very quickly to a rank-1 operator as m increases. This means that for m large enough so that $1/2^m$ becomes negligible, x_m, y_m, z_m are proportional to $\mathcal{E}_1^m$:

$$\mathbb{E}_{\{\mathcal{G}_i\}} S(\{\mathcal{G}_i\}) \approx \mathrm{diag}(1, \mathcal{C}_1 \mathcal{E}_1^m, \mathcal{C}_2 \mathcal{E}_1^m, \mathcal{C}_3 \mathcal{E}_1^m) \ , \tag{5.30}$$

where $\mathcal{C}_i$ are proportionality constants. Equation (5.3) is obtained by incorporating the SPAM procedures in evaluating the recovery probabilities, and by relabeling $\mathcal{E}_1$ as $p_{\mathbb{N}}$.

Finally, the relation between the decay $p_{\mathbb{N}}$ and the gate-set infidelity $r_{\mathbb{N}} = r(\mathcal{E}, \mathbb{I})$ is retrieved via eq. (5.29a):

$$r_{\mathbb{N}} = (1 - p_{\mathbb{N}})/2 + O(r_N^2) \ , \tag{5.31}$$

which essentially states that the NIST RB decay parameter $p_{\mathbb{N}}$ provides a very good estimates of the gate-set infidelity $r_{\mathbb{N}}$ through eq. (5.4).

With this analysis behind us, let's compare the internal mechanics of protocols 5 and 6. First of all, both protocols make use of randomizing gate-sets, $\mathbb{C}$ and $\mathbb{N}$ respectively. In both cases, the randomization tailors the error dynamics such that the average core sequence $\mathbb{E}_{\{\mathcal{G}_i\}} S(\{\mathcal{G}_i\})$ evolves with respect to a single decay parameter, as show in eqs. (5.16) and (5.30).

An interesting difference here is that the Clifford randomization simplifies the error into a 1-parameter depolarizing channel at each time step, while the NIST randomization doesn't, as shown in eq. (5.22). In the latter case, certain error components remain "imperfectly shuffled" after a few random gates, leaving space for a multi-parameterized noise evolution portrayed by eqs. (5.27) and (5.28). However, as the random sequence gets longer, the evolution quickly converges to a 1-parameter decay. The fact that this decay relates to the infidelity shouldn't be surprising, since $\mathrm{diag}(0, 1, 1, 1)$ is a channel component that commutes with every unitary (so is immune to twirling). Given an error $\mathcal{E}$, its corresponding coefficient is $(\mathcal{E}_{22} + \mathcal{E}_{33} + \mathcal{E}_{44})/3$, which has a one-to-one correspondence with the infidelity $r(\mathcal{E}, \mathbb{I})$ via eq. (5.17).

5.4.3 Measured Error Rates under NIST RB vs SRB

In the previous section, we showed that under the assumption of gate-independent errors, the gate-set infidelities $r_{\mathbb{C}}$ and $r_{\mathbb{N}}$ could be estimated via SRB and NIST RB respectively, and that these estimates did both coincide with $r(\mathcal{E}, \mathbb{I})$. In reality, one might find through experiment that $r_{\mathbb{C}}$ and $r_{\mathbb{N}}$ differ quite substantially. This, of course, is explained by gate-dependent effects: certain gates have higher infidelities than others, and since $r_{\mathbb{C}}$ and $r_{\mathbb{N}}$ each describe the expected gate infidelity *over their respective gate-set* (see section 5.6.1 for further justification), they will yield different values.

To demonstrate this, we numerically simulated both SRB and NIST RB experiments implemented in different fashions, using a plethora of primitive pulse sets (see table 5.1) each undergoing various physically realistic gate-dependent noise scenarios. The infidelities retrieved from each protocol $r_{\mathbb{C}}$ and $r_{\mathbb{N}}$ are juxtaposed in fig. 5.3, and differ by up to a factor

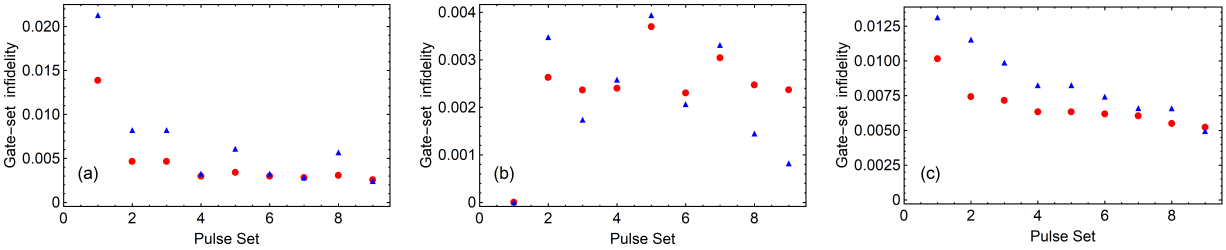


Figure 5.3: (Color online) A comparison between the gate-set infidelities $r_{\mathbb{N}}$ (blue triangles) and $r_{\mathbb{C}}$ (red points) for pulse sets from table 5.1 with gate-dependent (a) coherent over-rotation error $\tilde{\mathbb{I}} = \mathbb{I}$, $\tilde{X}_{\pm\theta} = X_{\pm(\theta+0.1)}$, $\tilde{Y}_{\pm\theta} = Y_{\pm(\theta+0.1)}$, $\tilde{Z}_{\pm\theta} = Z_{\pm(\theta+0.1)}$ (b) coherent Z-rotation error $\tilde{\mathbb{I}} = Z_{0.1}$, $\tilde{X}_{\theta} = Z_{0.1}X_{\theta}$, $\tilde{Y}_{\theta} = Z_{0.1}Y_{\theta}$, $\tilde{Z}_{\theta} = Z_{\theta+0.1}$, and (c) incoherent dephasing error $\tilde{\mathbb{I}} = D_{0.99}$, $\tilde{X}_{\theta} = D_{0.99}X_{\theta}$, $\tilde{Y}_{\theta} = D_{0.99}Y_{\theta}$, $\tilde{Z}_{\theta} = D_{0.99}Z_{\theta}$, where $D_{\alpha} = \mathrm{diag}(1, \alpha, \alpha, 1)$. Under these error models and pulse sets, $r_{\mathbb{N}}$ and $r_{\mathbb{C}}$ differ by up to a factor of 3, which could significantly affect the expected overhead under quantum error correction.

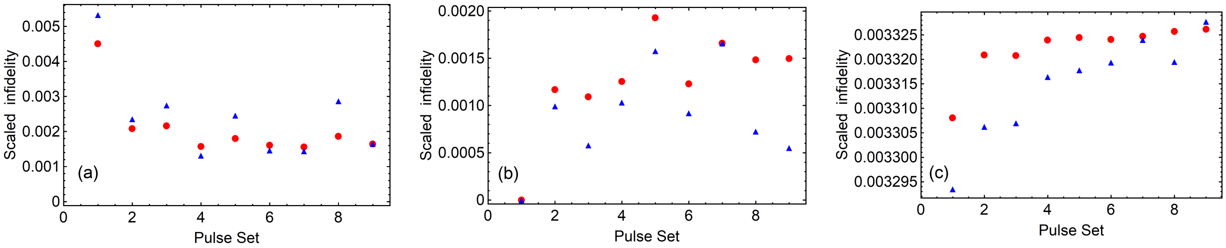


Figure 5.4: (Color online) A comparison between the scaled infidelities $r_{\mathbb{N}}/n_{\mathbb{N}}$ (blue triangles) and $r_{\mathbb{C}}/n_{\mathbb{C}}$ (red points) using pulse sets from table 5.1 with gate-dependent (a) coherent over-rotation error $\tilde{\mathbb{I}} = \mathbb{I}$, $\tilde{X}_{\pm\theta} = X_{\pm(\theta+0.1)}$, $\tilde{Y}_{\pm\theta} = Y_{\pm(\theta+0.1)}$, $\tilde{Z}_{\pm\theta} = Z_{\pm(\theta+0.1)}$ (b) coherent Z-rotation error $\tilde{\mathbb{I}} = Z_{0.1}$, $\tilde{X}_{\theta} = Z_{0.1}X_{\theta}$, $\tilde{Y}_{\theta} = Z_{0.1}Y_{\theta}$, $\tilde{Z}_{\theta} = Z_{\theta+0.1}$, and (c) incoherent dephasing error $\tilde{\mathbb{I}} = D_{0.99}$, $\tilde{X}_{\theta} = D_{0.99}X_{\theta}$, $\tilde{Y}_{\theta} = D_{0.99}Y_{\theta}$, $\tilde{Z}_{\theta} = D_{0.99}Z_{\theta}$, where $D_{\alpha} = \mathrm{diag}(1, \alpha, \alpha, 1)$. Clearly, even after accounting for the discrepancy between the average number of pulses per gate of these two RB protocols ($n_{\mathbb{N}}$ & $n_{\mathbb{C}}$), the measured error rates still differ due to their differing sampling over the gate sets.

Index	Pulse Set	$n_{\mathbb{C}}$	$n_{\mathbb{N}}$
1	$\{\mathbb{I}, \tilde{X}_{+\pi/2}, \tilde{Y}_{+\pi/2}\}$	3.08333	4.0
2	$\{\tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	2.25	3.5
3	$\{\mathbb{I}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	2.16667	3.0
4	$\{\tilde{X}_{\pi}, \tilde{Y}_{\pi}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	1.91667	2.5
5	$\{\tilde{\mathbb{I}}, \tilde{Z}_{\pi}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	1.91667	2.5
6	$\{\tilde{\mathbb{I}}, \tilde{X}_{\pi}, \tilde{Y}_{\pi}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	1.875	2.25
7	$\{\tilde{\mathbb{I}}, \tilde{X}_{\pi}, \tilde{Y}_{\pi}, \tilde{Z}_{\pi}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	1.8333	2.0
8	$\{\mathbb{I}, Z_{\pi}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	1.66667	2.0
9	$\{\mathbb{I}, \tilde{X}_{\pi}, \tilde{Y}_{\pi}, Z_{\pi}, \tilde{X}_{\pm\pi/2}, \tilde{Y}_{\pm\pi/2}\}$	1.58333	1.5

Table 5.1: Each of the 24 $\mathbb{C}$s and 8 $\mathbb{N}$s were constructed by a sequence of noisy ($\tilde{\mathbb{I}}, \tilde{X}_{\theta}, \tilde{Y}_{\theta}, \tilde{Z}_{\theta}$) and virtually (ideal) implemented ($\mathbb{I}, X_{\theta}, Y_{\theta}, Z_{\theta}$) pulses. Note that when implementing the π pulses, the direction of rotation (sign of π) is selected uniformly at random as described in [92]. $n_{\mathbb{C}}$ and $n_{\mathbb{N}}$ are the average number of noisy pulses per gate from $\mathbb{C}$ and $\mathbb{N}$, and is used for calculating the scaled infidelity ($r_{\mathbb{C}}/n_{\mathbb{C}}$ and $r_{\mathbb{N}}/n_{\mathbb{N}}$).

of ~ 3. This might not strike as a major difference in a day and age where the infidelity is typically filtered through its order of magnitude. That being said, for both surface codes and concatenated quantum error correcting codes, the overhead becomes more sensitive to the error rate as it approaches the fault-tolerant threshold. Therefore, a factor of 3 could dramatically increase the overhead (by more than an order of magnitude) if the error rate is close to the threshold. In the extreme case, the factor of 3 could cause the error rate to surpass the threshold and fault tolerance would become impossible.

One might be tempted to reconcile those different infidelities by accounting for the number of primitive pulses that form each gate-set element. That is, if the Clifford gates $C \in \mathbb{C}$ necessitate the average application of $n_{\mathbb{C}} = 1.875$ pulses and that the NIST gates $N \in \mathbb{N}$ require an average of $n_{\mathbb{N}} = 2.25$ pulses, it might seem natural to scale the infidelities as $r_{\mathbb{C}}/n_{\mathbb{C}}$ and $r_{\mathbb{N}}/n_{\mathbb{N}}$. This is a common misconception because scaling the infidelity is only meaningful if it grows linearly with the number of pulses, as for a purely stochastic error. For example, in gate-dependent dephasing scenarios, the scaled infidelities ($r_{\mathbb{C}}/n_{\mathbb{C}}$ and $r_{\mathbb{N}}/n_{\mathbb{N}}$) only differ by $O(r_{\mathbb{N}}^2)$ because the error is purely stochastic (incoherent) and pulse-independent (see fig. 5.4c).

Contrarily, it has been shown that under coherent error scenarios, the composite errors can vary non-linearly [33], as they can positively and negatively interfere. Therefore, when the error is coherent, NIST RB and SRB obtain different scaled infidelities, despite the infidelity per pulse remaining fixed (see figs. 5.4a and 5.4b) because they are sampling differently from the pulse sets, which causes them to probe different coherent error models. As such, it is bad practice to measure the scaled infidelity which is not equivalent to, and should not be confused with, the error per pulse.

5.5 Conclusion

Randomized benchmarking (RB) is an important tool for estimating error rates associated with sets of elementary gate operations. SRB and NIST RB are two distinct RB protocols that have been confused in the literature and can lead to distinct outcomes. In this work we developed a rigorous theoretical framework proving that NIST RB, like SRB, leads to an exponential decay which depends only on the underlying gate-independent error model. We showed SRB, which samples from a uniform 2-design, and NIST RB, which samples from a subset of it, lead to significantly different observed error rates for a variety of physically realistic gate-dependent error models and pulse sets (see fig. 5.3). The common experimental error models that we use throughout this study, we obtain error rates which differ by up to a factor of 3, which could have a significant impact on the overhead when implementing fault-tolerant quantum error correction.

A next step is to develop a rigorous theoretical framework under which the experimental results from NIST RB and other RB methods using arbitrary gate sets can be analyzed in a unified and related to infer properties of the underlying error model in a consistent manner.

5.6 Supplementary Material

5.6.1 NIST RB analysis under gate-dependent noise

In this section, we broadly justify the validity of NIST RB in the case of gate-dependent error models. Since a complete analysis would necessitate pages of mathematical developments, we first direct the reader to [171, 31, 116, 133].

The essence of the proof of the decay model eq. (5.3) resides in realizing that the RB recovery probabilities evolve as a combination of decays $\sum_i p_i^m$, where p_i are the eigenvalues of $\mathbb{E}_{\mathbb{N}}\mathcal{G} \otimes \tilde{\mathcal{G}}$. Those eigenvalues are slightly perturbed from those of $\mathbb{E}_{\mathbb{N}}\mathcal{G} \otimes \mathcal{G}$. In the case of the NIST gate-set $\mathbb{N}$, the non-zero eigenvalues are $1, 1, 1/2$ and $-1/2$. In the perturbed case ($\mathbb{E}_{\mathbb{N}}\mathcal{G} \otimes \tilde{\mathcal{G}}$), the first eigenvalue remains 1, as the trace-preservation property causes it to remain invariant, and the second one becomes $p_{\mathbb{N}} \approx 1$. The two eigenvalues which are close to $\pm 1/2$ decay very fast as the circuit grows, and don't contribute to the decay model for large m.

The relationship between the decay constant $p_{\mathbb{N}}$ and the infidelity $r_{\mathbb{N}}$ (eq. (5.4)) can be derived as a straightforward generalization of the analysis derived in [31]. Let the eigenvector corresponding to the decay $p_{\mathbb{N}}$ be

$$\mathbb{E}_{\mathbb{N}}\mathcal{G} \otimes \tilde{\mathcal{G}}\mathrm{vec}(L) = p_{\mathbb{N}}\mathrm{vec}(L) \ , \tag{5.32}$$

where $\mathrm{vec}(\cdot)$ is the column vectorization. In a nutshell, eq. (5.4) holds as long as the singular values of L are close to each other, which is shown to be the case in 1-qubit SRB [31]. The reasoning, as it pertains to the NIST gate-set, goes as follows: Let Π_{tr} be the 3×3 projector on the traceless hyperplane (the Bloch space). Given the spectrum of $\mathbb{E}_{\mathbb{N}}\mathcal{G} \otimes \tilde{\mathcal{G}}$, we have:

$$L \approx \mathbb{E}\left(\frac{\tilde{\mathcal{G}}_m\tilde{\mathcal{G}}_{m-1}\cdots\tilde{\mathcal{G}}_1\mathcal{G}_{m:1}^{-1}}{p_{\mathbb{N}}^m}\right)\Pi_{\mathrm{tr}} \ , \tag{5.33}$$

for m large enough so the r.h.s converges. Indeed, performing $p_{\mathbb{N}}^{-1}\mathbb{E}_{\mathbb{N}}\Pi_{\text{tr}}\mathcal{G}\otimes\tilde{\mathcal{G}}$ multiple times converges very quickly to a rank-1 projector onto the desired eigenspace. Since L is the result of a reasonably short sequence of noisy operations (say $m = -2\log(r)/\log(2)$), it is proportional to a high-fidelity channel, for which the singular values are close to each other (at least in the single-qubit case).

Chapter 6

Gold standard for cross-platform comparisons of basic gate-sets

6.1 Overview

Quantum computers will have the potential to solve some of humanity's greatest problems. One of the main barriers to the development of quantum computers is that quantum errors are ubiquitously present in quantum systems. We propose a set of optimal protocols to assess the performance of single- and two-qubit gates, both for internal use to optimize device performance as well as to make cross-platform comparisons. The former use case provides insight into device performance for specific applications, allowing a device to be tuned effectively for a given use-case by focusing on the relevant gate-set. This enables targeted tune-up of control pulses, optimization of pulse-design, and automation of calibration, which will reduce error and thus will allow hardware developers to increase the complexity of their quantum computations. The cross-platform comparisons of different primitive universal gate-sets includes an even-handed assessment of any two-qubit entangling gate, and rigorous systematic error bounds to estimate the worst case error. We report results from implementing these protocols across a variety of leading platforms, including both superconducting qubits and trapped ions.

6.2 Introduction

To develop quantum devices capable of solving real world problems, the system size and sequence length of operations must continue to grow. Unfortunately, due to the cross-talk between qubits, the amount of error grows exponentially with the system size, and grows non-linearly with sequence length as coherent errors generally positively interfere with each other. In other words, quantum errors are the Achilles heel to the development of a universal quantum computer. Error characterization schemes will play an important role in both the NISQ era and in the long run because they help guide the improvement of devices by aiding in calibration, and making various assessments (FTQC or NISQ) to test the reliability of computations. There are various schemes for analyzing the performance of quantum devices [101, 83] ranging from tomographic approaches [41, 129] to more “streamlined” protocols such as Randomized Benchmarking[49, 95, 106]. The abundance of possible diagnostic tools leads us to several important questions: Which protocols should be used to evaluate the noise in a quantum device, and in which context? How do we combine them? What kind of diagnoses can we make from the data?

In this chapter, we propose the basic gate-set assessment (BGSA), a well-rounded standard for characterizing single- and 2-qubit errors which can be used for any gate-based quantum device and entangling gate. We argue that the BGSA can be used to make an even-handed, cross-platform comparison of universal gate-sets. Since each platform uses unique techniques to implement quantum gates, a cross-platform comparison must focus on assessing gates rather than evaluating the hardware-level pulses directly. The BGSA can also be used for local tune-up, and if the amount of cross-talk is low, then it is enough to tune up large quantum devices each pair at a time.

Prior to this work, an even-handed cross-platform comparison had not been proposed due to the many device-specific restrictions, the many degrees of freedoms when running benchmarking protocols, and the uncertainty about how to compare results from such different implementations. It was previously unclear how to interpret Randomized Benchmarking data when gate depths varied significantly within a gate-set. Since benchmarking protocols usually provide fidelity metrics for entire gate-sets, the gate-set would need to be chosen such that it could be unbiasedly implemented by any device. Since different platforms use different elementary gates, they each have very different gate-set implementations, and even more alarming, the fidelity may vary greatly from one gate-set element to the other on the same device. Additionally, different gate-sets may be used for different purposes by the same device, and so it is important to choose a gate-set which will provide an unbiased comparison. For example the use of the SU(4) group is great for comparing how well a device can implement a universal gate-set, while the Clifford group is better when trying to recalibrate the system due to its more standard gates.

Many 2-qubit gates which are seen in literature fall within the 2-qubit Clifford group (examples include CNOT, CZ, iSWAP, and SWAP). All 2-qubit Cliffords can be subdivided into 4 distinct sets: 1) tensor-like (i.e. the tensor product of 2 single-qubit Cliffords), 2) CNOT-like, 3) iSWAP-like, and 4) SWAP-like[11]. These 4 sets can be thought of as equivalence classes up to single-qubit Cliffords, meaning that any gate inside a set is equivalent to all other gates inside that set up to conjugation by single-qubit Clifford operations. Most physical implementations of 2-qubit gates typically attempt to implement either a CNOT-like or iSWAP-like operation. Since some devices implement a 2-qubit gate with a single 2-qubit pulse and others with a 2-qubit pulse and multiple single-qubit pulses, the depth of the implementation can vary between devices even when applying the same gate. Sampling a gate at random from the 2-qubit Cliffords can therefore lead to variations in total pulse count depending on what physical gates are available on hardware.[1] This problem is exacerbated when hardware does not use Clifford-like operations [7] in their elementary gate-set. In those situations, the composition of physical gates may be close to Clifford gates, but the imperfections mean the decomposition into several pulses may lead to large variations in gate depths.

Sampling from SU(4), however, almost uniformly leads to the same total gate depth no matter which element of SU(4) is chosen. A randomly chosen element of SU(4) has a vanishingly small probability of being in $\mathbb{C}_2$, and as a result any CNOT-like or iSWAP-like gates will require exactly 3 entangling gates plus single-qubit operations to decompose precisely. Numerically, we demonstrate that this property of entangling gate depth extends beyond those two Clifford groups to gates which are "close" to being Clifford as well, so that any gate which is sufficiently "close" to a Clifford gates can be decomposed into a form requiring exactly 3 entangling gates. The gates in ref [7] satisfy this condition. There are a number of applications where it is necessary to perform operations which are non-Clifford. VQE (Variational-Quantum-Eigensolver) is one example of such an application, which is a hybrid classical and quantum algorithm to find the eigenvalues of a large matrix. In such a situation, hardware which is unable to perform a native Clifford gate can still perform as well as hardware which has native Clifford operations.

To ensure that gate depth does not impact performance metrics, when the BGSA is used to make an unbiased, cross-platform comparison of the error acting on a universal gate-set, the gates are sampled from SU(4). Conversely, when the BGSA is used internally for local performance optimization, we recommend sampling from the 2-qubit Clifford group. While

---

[1] Note that this fluctuation in the number of gates can be reduced for 2-qubit Clifford RB by compiling into a fixed depth.

the Clifford group is not a universal gate-set, we recommend using $\mathbb{C}_2$ as its elements can be implemented with circuits of lower depth than those of SU(4). For very high error rates, the fidelity will decay less quickly for $\mathbb{C}_2$ than when using SU(4), and therefore the estimate of the single gate process infidelity of the entangling gate will be more stable.

Process infidelity (e_F) is the figure of merit used in this chapter to characterize various aspects of the error model. For any channel A and B

$$1 - e_F(A \otimes B) = (1 - e_F(A))(1 - e_F(B)), \tag{6.1}$$

or

$$e_F(A \otimes B) = e_F(A) + e_F(B) - e_F(A)e_F(B) \approx e_F(A) + e_F(B), \tag{6.2}$$

This relationship does not hold for the more commonly used average gate infidelity, (r). The conversion between the process infidelity and average gate infidelity is:

$$e_F = \frac{d+1}{d} r. \tag{6.3}$$

The process infidelity is related to the depolarizing parameter by:

$$p = 1 - \frac{d^2 e_F}{d^2 - 1}. \tag{6.4}$$

The stochastic process infidelity (e_S) from XRB can be written in terms of the unitarity (u) as:

$$e_S = 1 - \sqrt{\frac{(d^2-1)u + 1}{d^2}} \tag{6.5}$$

Section 6.3 is a summary of the BGSA. This includes an overview of how each protocol is implemented, the physical interpretation of the returned figures of merit, how these figures of merit can be used together to give a well-rounded picture of the error profile, the twirling set used for each protocol, the number of circuits needed to implement each protocol, and a summary of our recommendations for the various user freedoms.

Section 6.4 discusses the experimental results obtained by implementing the BGSA on two superconducting devices (Yorktown$_5$[2] a 5 qubit device from IBM and AQT_4 an 8 qubit device with 4 operational qubits at the time of implementation from LBNL), and one ion-trap device (H0$_6$ a 6 qubit device from Honeywell), making it the first even-handed

---

[2] The notation here is DevName$_{nqubits}$ where Devname is the name of the device, and nqubits is the total number of operational qubits the device had at the time of implementation.

cross-platform comparison of quantum devices. Note that the results obtained from the Yorktown device were from late 2019 on one of IBM's open devices, and therefore do not showcase how well their premium devices currently operate.

We begin by first examining the single-qubit gates of these devices and the amount of single-qubit cross-talk there is between the target pair of qubits. Then we discuss the various error estimates associated with the two-qubit gates. This includes the amount of coherent and stochastic error effecting each device, how the dressed gate process infidelity from IRB compares to the bare gate process infidelity from SRB, how well those values are used together to estimate the single gate process infidelity of the entangling gate, how the systematic upper bound can be tightened significantly by accounting for the amount of stochastic error, how the coherent upper bound from IRB compares to the systematic upper bound from CB, and when using CB, how well the dressed gate process infidelity approximates the single gate estimate when the $\mathbb{P}_2$ error is insignificant.

6.3 Basic Gate-Set Assessment

In this section, we propose a simple, fast, and non-controversial suite of benchmarking protocols that assesses the performance of single- and 2-qubit gates. The suite, referred to as the Basic Gate-Set Assessment (BGSA), includes four types of experiments, namely Standard Randomized Benchmarking (SRB), Interleaved Randomized Benchmarking (IRB), Extended Randomized Benchmarking (XRB), and Cycle Benchmarking (CB). For historical continuity, the single-qubit assessment is done with a Clifford-based SRB.

The choice of gate set used in the assessment depends on the underlying goal. When the objective is to participate in an unbiased cross-platform comparison, we recommend running the BGSA on many – if not all – connected qubit pairs in order to find the best pair because the same gate performed on different qubits will have a different noise profile. We therefore benchmark the best pair of qubits for each device to obtain a fair comparison. The twirling group used for the 2-qubit assessment must be SU(4). The use of SU(4) allows the BGSA to benchmark a universal gate set (universal gatesets are primitives for quantum volume [44]), and make an unbiased comparison between arbitrary native entangling gates ($\mathcal{G}$) with different entangling power which enables a comparison of platforms with and without native Clifford gates.

It is common to compare platforms by comparing the fidelities of the respective entangling operations[83]. Some devices may have a high-fidelity entangling gate, but require many gates to generate an element of SU(4), while a different device may have a lower-fidelity entangling gate, but require fewer of those gates on average to generate an element of SU(4). This is a motivation behind defining the quantum volume as a cross-platform metric[44]. Hence, for the sake of a cross-platform comparison, it is best to compare SU(4) fidelities rather than the fidelities of entangling gates, which may not be operationally equivalent.

Calibration rounds mostly target individual gates of interest, in which case IRB-like experiments become central. There is no particular need to reduce the variation of gate depths within the randomizing set. Rather, it is beneficial to choose a randomizing set that has the highest fidelity. Higher fidelity randomizing sets provide higher resolution estimates for single gate fidelity[85]. Compared to $\mathbb{C}_2$, SU(4) exhibits less variation in the gate depth required to implement a given element, but generally has a lower fidelity. SU(4) has greater circuit depth and complexity for each of the RB operations, so for very high error rates, the fidelity decays too fast which causes the IRB estimate to be subject to high precision loss.

In this section we will describe the protocols using SU(4) instead of $\mathbb{C}_2$, for simplicity. In practice, users can choose to use either SU(4) or $\mathbb{C}_2$ to implement BGSA, depending on their use-case. When the purpose of implementing the BGSA is for local tune-up, use the suite of protocols from table 6.1, and when used to participate in a cross-platform comparison, use the protocols from table 6.2. Note that the 2-qubit Paulis ($\mathbb{P}_2$) are used in the Cycle Benchmarking experiments. Three sequence lengths are taken for all SRB protocols to test the Markovianity of the error.

In the following two subsections, we list all of the protocols contained in BGSA and what information they retrieve about the noise profile. We argue that in combination, this set of protocols provides a well-rounded understanding of the error profile, while keeping the number of circuits needed to a reasonably low count.

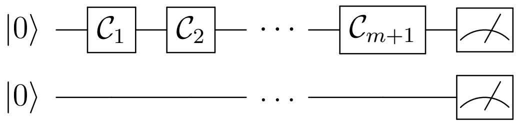
6.3.1 Single-qubit gates benchmarking experiments

Isolated SRB

Using the single-qubit Clifford group ($\mathbb{C}_1$) as the twirling group, we calculate the isolated single-qubit process infidelities $e_F(\mathbb{C}_{q_1}|\mathbb{C}_{q_1} \otimes \mathbb{I}_{q_2})$ and $e_F(\mathbb{C}_{q_2}|\mathbb{I}_{q_1} \otimes \mathbb{C}_{q_2})$ of each of the two qubits of interest via SRB (from protocol 1). This figure of merit is the average error induced by single-qubit operations, when they are applied in isolation.

Protocol	Estimate	Twirling Group	Interleaved/ Hard Gate	# Circuits
Isolated SRB	$e_F(\mathbb{C}_{q_1}\vert\mathbb{C}_{q_1}\otimes\mathbb{I}_{q_2})$	$(\mathbb{C}_1,q_1)$		90
	$e_F(\mathbb{C}_{q_2}\vert\mathbb{I}_{q_1}\otimes\mathbb{C}_{q_2})$	$(\mathbb{C}_1,q_2)$		90
Simultaneous SRB	$e_F(\mathbb{C}_{q_1}\vert\mathbb{C}_{q_1}\otimes\mathbb{C}_{q_2})$	$((\mathbb{C}_1,q_1),(\mathbb{C}_1,q_2))$		90
	$e_F(\mathbb{C}_{q_2}\vert\mathbb{C}_{q_1}\otimes\mathbb{C}_{q_2})$	$((\mathbb{C}_1,q_1),(\mathbb{C}_1,q_2))$		90
SRB	$e_F(\mathbb{C}_2)$	$(\mathbb{C}_2,q_1,q_2)$		90
XRB	$e_S(\mathbb{C}_2)$	$(\mathbb{C}_2,q_1,q_2)$		540
IRB	$e_F(\mathbb{C}_2\mathcal{G})$	$(\mathbb{C}_2,q_1,q_2)$	$\mathcal{G}$	60
CB	$e_F(\mathbb{P}_2)$	$(\mathbb{P}_2,q_1,q_2)$	$I\otimes I$	540
	$e_F(\mathbb{P}_2\mathcal{G})$	$(\mathbb{P}_2,q_1,q_2)$	$\mathcal{G}$	540

Table 6.1: The protocols which make up the BGSA when used for calibration, along with their corresponding estimated figures of merit, twirling group, interleaved gate (where $\mathcal{G}$ is the entangling gate of the device), and number of different circuits used for the implementation of the protocols. The notation for a group acting on an n-qubit system $\mathbb{C}_n$, should not be confused for the notation for a single-qubit group acting on qubit a, $\mathbb{C}_{q_a}$. The process infidelity of $\mathbb{C}_{q_1}$ ($\mathbb{C}_{q_2}$) is given by $e_F(\mathbb{C}_{q_1}|\mathbb{C}_{q_1}\otimes\mathbb{I}_{q_2})$ $(e_F(\mathbb{C}_{q_2}|\mathbb{I}_{q_1}\otimes\mathbb{C}_{q_2}))$[3]when single-qubit gates are run in isolation, and $e_F(\mathbb{C}_{q_1}|\mathbb{C}_{q_1}\otimes\mathbb{C}_{q_2})$ $(e_F(\mathbb{C}_{q_2}|\mathbb{C}_{q_1}\otimes\mathbb{C}_{q_2}))$ when they are run simultaneously. The process infidelity of $\mathbb{C}_2$ ($\mathbb{P}_2$) is given by $e_F(\mathbb{C}_2)$ $(e_F(\mathbb{P}_2\mathcal{G}))$, and its stochastic process infidelity is given by $e_S(\mathbb{C}_2)$, while the process infidelity of the dressed cycle is given by $e_F(\mathbb{G}\mathcal{G})$ $(e_F(\mathbb{P}_2))$. See table 6.4 for experimental results.



For $\mathcal{C}_i \in \mathbb{C}_1$ and $\mathcal{C}_{m+1} = \mathcal{C}^{-1}_{\{1:m\}} = (\mathcal{C}_m \circ \cdots \circ \mathcal{C}_1)^{-1}$

Simultaneous SRB

Twirling with the single-qubit Clifford group, we calculate the simultaneous single-qubit process infidelities $e_F(\mathbb{C}_{q_1}|\mathbb{C}_{q_1}\otimes\mathbb{C}_{q_2})$ and $e_F(\mathbb{C}_{q_2}|\mathbb{C}_{q_1}\otimes\mathbb{C}_{q_2})$ on each of the two qubits of

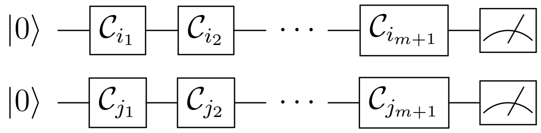
---

[3] The notation $e_F(A|A\otimes B)$, is the process infidelity of $Tr_B(A\otimes B)$, to obtain the process infidelity of just A (see eq. (2.21)).

Protocol	Estimate	Twirling Group	Interleaved/ Hard Gate	# Circuits
Isolated SRB	$e_F(\mathbb{C}_{q_1}\|\mathbb{C}_{q_1} \otimes \mathbb{I}_{q_2})$	$(\mathbb{C}_1,q_1)$		90
	$e_F(\mathbb{C}_{q_2}\|\mathbb{I}_{q_1} \otimes \mathbb{C}_{q_2})$	$(\mathbb{C}_1,q_2)$		90
Simultaneous SRB	$e_F(\mathbb{C}_{q_1}\|\mathbb{C}_{q_1} \otimes \mathbb{C}_{q_2})$	$((\mathbb{C}_1,q_1),(\mathbb{C}_1,q_2))$		90
	$e_F(\mathbb{C}_{q_2}\|\mathbb{C}_{q_1} \otimes \mathbb{C}_{q_2})$	$((\mathbb{C}_1,q_1),(\mathbb{C}_1,q_2))$		90
SRB	$e_F(SU(4))$	(SU(4),q_1,q_2)		90
XRB	$e_S(SU(4))$	(SU(4),q_1,q_2)		540
IRB	$e_F(SU(4)\mathcal{G})$	(SU(4),q_1,q_2)	$\mathcal{G}$	60
CB	$e_F(\mathbb{P}_2)$	$(\mathbb{P}_2,q_1,q_2)$	$I \otimes I$	540
	$e_F(\mathbb{P}_2\mathcal{G})$	$(\mathbb{P}_2,q_1,q_2)$	$\mathcal{G}$	540

Table 6.2: The protocols which make up the BGSA cross-platform comparison, along with their corresponding estimated figures of merit, twirling group, interleaved gate (where $\mathcal{G}$ is the entangling gate of the device), and number of different circuits used for the implementation of the protocols. The process infidelity of $\mathbb{C}_{q_1}$ ($\mathbb{C}_{q_2}$) is given by $e_F(\mathbb{C}_{q_1}|\mathbb{C}_{q_1} \otimes \mathbb{I}_{q_2})$ ($e_F(\mathbb{C}_{q_2}|\mathbb{I}_{q_1} \otimes \mathbb{C}_{q_2})$) when single-qubit gates are run in isolation, and $e_F(\mathbb{C}_{q_1}|\mathbb{C}_{q_1} \otimes \mathbb{C}_{q_2})$ ($e_F(\mathbb{C}_{q_2}|\mathbb{C}_{q_1} \otimes \mathbb{C}_{q_2})$) when they are run simultaneously. The process infidelity of SU(4) ($\mathbb{P}_2$) is given by $e_F(SU(4))$ ($e_F(\mathbb{P}_2\mathcal{G})$), and its stochastic process infidelity is given by $e_S(SU(4))$, while the process infidelity of the dressed cycle is given by $e_F(SU(4)\mathcal{G})$ ($e_F(\mathbb{P}_2)$). See table 6.4 for experimental results.

interest by running single-qubit SRB (protocol 1) simultaneously on each qubit.



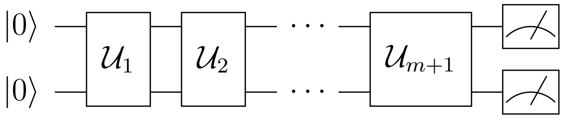
For $\mathcal{C}_{i_k}, \mathcal{C}_{j_k} \in \mathbb{C}_1$ and $\mathcal{C}_{i_{m+1}} \otimes \mathcal{C}_{j_{m+1}} = \mathcal{C}^{-1}_{\{1:m\}} = (\mathcal{C}_{i_m} \circ \cdots \circ \mathcal{C}_{i_1})^{-1} \otimes (\mathcal{C}_{j_m} \circ \cdots \circ \mathcal{C}_{j_1})^{-1}$

The difference between the simultaneous single-qubit process infidelity and the isolated single-qubit process infidelity is the process infidelity of a type of cross-talk between the two qubits.

6.3.2 2-qubit gates benchmarking experiments

SRB

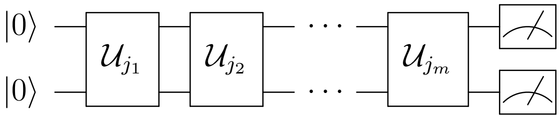
Isolated 2-qubit SRB (from protocol 1) tests the quality of the 2-qubit registers by calculating the *bare gate process infidelity*, $e_F(SU(4))$, of the SU(4) group.



For $\mathcal{U}_i \in \mathrm{SU}(4)$ and $\mathcal{U}_{m+1} = \mathcal{U}^{-1}_{\{1:m\}} = (\mathcal{U}_m \circ \cdots \circ \mathcal{U}_1)^{-1}$.

XRB

XRB (protocol 3) produces an exponential fidelity decay which is used to calculate the the stochastic process infidelity of the error, $e_S(SU(4))$, which is related to the unitarity via eq. (6.5).



e_S can be used to calculate the coherent process infidelity e_U [30]:

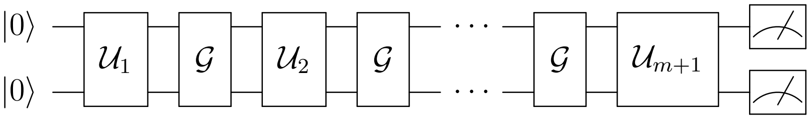
$$e_U(SU(4)) = e_F(SU(4)) - e_S(SU(4)). \tag{6.6}$$

e_U indicates the maximum amount of error that could be caused by correctable miscalibrations of the system. If this value is large, than it suggests that some improvement can be made via re-calibration.

The stochastic process infidelity can also be used to tighten the systematic error bound of the single gate process infidelity, $e_F(\mathcal{G})$ [34].

IRB

Isolated 2 qubit IRB (protocol 2) is used to calculate the *dressed gate process infidelity*, $e_F(SU(4)\mathcal{G})$, of SU(4) composed with the chosen entangling gate $\mathcal{G}$,



where $\mathcal{U}_{m+1}$ is the inverse of the total sequence.

The *single gate process infidelity* $e_F(\mathcal{G})$ of the entangling gate $\mathcal{G}$ can be approximated by combining the data retrieved by isolated 2-qubit IRB with the previously generated SU(4) SRB data via

$$e_F(\mathcal{G})_{SU(4)} \approx e_F(SU(4)\mathcal{G}) - e_F(SU(4)), \tag{6.7}$$

which is bounded by the *systematic bound* from [34]

$$\begin{aligned} &\left|1 - e_F(\mathcal{G})_{SU(4)} - e_F(SU(4)\mathcal{G})e_F(SU(4)) - (1 - e_F(SU(4)\mathcal{G}))(1 - e_F(SU(4)))\right| \\ &\qquad \leq 2\sqrt{e_F(SU(4)\mathcal{G})e_F(SU(4))(1 - e_F(SU(4)\mathcal{G}))(1 - e_F(SU(4)))}. \end{aligned} \tag{6.8}$$

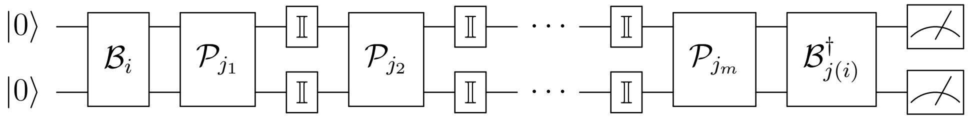
These systematic bounds can be tightened by taking into consideration the unitarity of the error profile [34]. For clarity, we refer to this tighter bound as the *coherent bound*.

$$\left|p(\mathcal{G})_{SU(4)} - \frac{p(SU(4)\mathcal{G})p(SU(4))}{u(SU(4))}\right| \leq \sqrt{1 - \frac{p(SU(4))^2}{u(SU(4))}}\sqrt{1 - \frac{p(SU(4)\mathcal{G})^2}{u(SU(4))}}, \tag{6.9}$$

where $p(SU(4))$ is the average gate depolarizing parameter of SU(4) obtained via SRB (see eq. (6.4)), $p(SU(4)\mathcal{G})$ is the average gate depolarizing parameter of the composition of the error from the entangling gate $\mathcal{G}$ with SU(4) obtained via IRB, and $u(SU(4))$ is the unitarity (see eq. (2.58)) of SU(4). This coherent bound is tighter for smaller unitarity. Therefore the closer the error is to stochastic, the tighter the coherent bound.

Bare gate CB

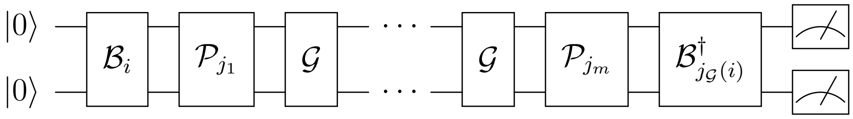
Similar to SRB, the *bare gate process infidelity* of the Pauli group, $e_F(\mathbb{P}_2)$, is calculated using Cycle Benchmarking (protocol 4) with $\mathbb{P}_2$ as the twirling group.



for $\mathcal{P} \in \mathbb{P}_2$. Note that here the hard/interleaved gate is the single-qubit $\mathbb{I}$, which can be removed when the circuits are compiled, and should not be physically implemented in practice.

Dressed gate CB

Similar to IRB, dressed gate CB (protocol 4) is used to calculate the average *dressed gate process infidelity* of the entangling gate ($\mathcal{G}$) composed with $\mathbb{P}_2$, $e_F(\mathbb{P}_2\mathcal{G})$.



These circuits are generated such that the m's differ by k such that $\mathcal{G}^k$ is factorizable.

The *single gate process infidelity* $e_F(\mathcal{G})_{\mathbb{P}_2}$ of the entangling gate can be calculated using the results from dressed gate CB and bare gate CB via:

$$e_F(\mathcal{G})_{\mathbb{P}_2} = e_F(\mathbb{P}_2\mathcal{G}) - e_F(\mathbb{P}_2) \approx e_F(\mathbb{P}_2\mathcal{G}). \tag{6.10}$$

As we will see in the next section, CB gives an estimate of the single gate process infidelity with significantly smaller statistical and systematic uncertainty calculated via [34]

$$\begin{aligned}|1 - e_F(\mathcal{G})_{\mathbb{P}_2} - e_F(\mathbb{P}_2\mathcal{G})e_F(\mathbb{P}_2) - (1 - e_F(\mathbb{P}_2\mathcal{G}))(1 - e_F(\mathbb{P}_2))| \\ \leq 2\sqrt{e_F(\mathbb{P}_2\mathcal{G})e_F(\mathbb{P}_2)(1 - e_F(\mathbb{P}_2\mathcal{G}))(1 - e_F(\mathbb{P}_2))},\end{aligned} \tag{6.11}$$

due to the low Pauli gate error.

6.3.3 BGSA user freedoms

The final choice is how to calculate the single gate process infidelity of the entangling gate. As has been presented above, the options are between running IRB and/or CB. If CB is

Freedom	Recommendation
Single-qubit Gate-set	Best
Entangling Gate	Best
Synthesize Operations	Best
Pair of Qubits	Best
Sequence Lengths	$[4, (1-p)^{-1}]$
Circuits/ Sequence Length	30
# of Shots	100
Order Protocols Run	See table 6.1 and 6.2
How to estimate $e_F(\mathcal{G})$	See Below

Table 6.3: User Freedoms and Recommendations. For the purpose of making a cross-platform comparison, we recommend running the BGSA on all possible pairs of qubits, entangling gates, single-qubit gate-sets, and ways to synthesize operations, such that the optimal construction can be used in order to showcase the full ability of the device. The other freedoms, the user should follow our recommendations, but in practise these freedoms can be adjusted such that the best fit can be made of the data (see section 6.6.1).

run then the remaining option is between running only dressed gate CB or running both dressed and bare gate CB. Here we recommend using the dressed gate process infidelity if the pauli error is insignificant in comparison. If IRB is run, and the user is running BGSA for internal tuning and calibration, then we recommend running IRB, XRB, and SRB using $\mathbb{C}_2$ instead of SU(4), as has been described above. If the user can implement the circuits sufficiently fast, then CB is recommended as it obtains the estimate with the smallest systematic and statistical uncertainty. In addition, if the error rate of the Pauli gates is insignificant relative to the entangling gate, then it is not necessary to run the bare gate CB as the single gate process infidelity of the entangling gate will be approximately equal to the process infidelity of the dressed gate found via CB. As will be described in the following section, in general the dressed gate infidelity via CB is a more meaningful estimate in practise anyways.

6.4 Interpreting the results

6.4.1 Single-qubit results

The infidelity of operations on a single-qubit generally depends on what happens in its surroundings. If two qubits are near each other, then it is hard to act on one without affecting the other. This contingent event is categorized as a type of cross-talk (see section 2.9.2). To quantify its significance, we recommend benchmarking single-qubit gates in two different contexts: first, when no operations are applied to nearby qubits, which corresponds to what we refer to as “isolated SRB”, and second, when single-qubit gates are applied to neighbouring qubits, which corresponds to what we refer to as “simultaneous SRB”. The difference between the fidelity estimates from isolated SRB and simultaneous SRB quantifies the aforementioned cross-talk effect. For example, in fig. 6.1, qubit 5 on the AQT_4 device has a smaller cross-talk error from qubit 6 than from qubit 4. Note that all of the estimates for the studied devices are given in table 6.4.

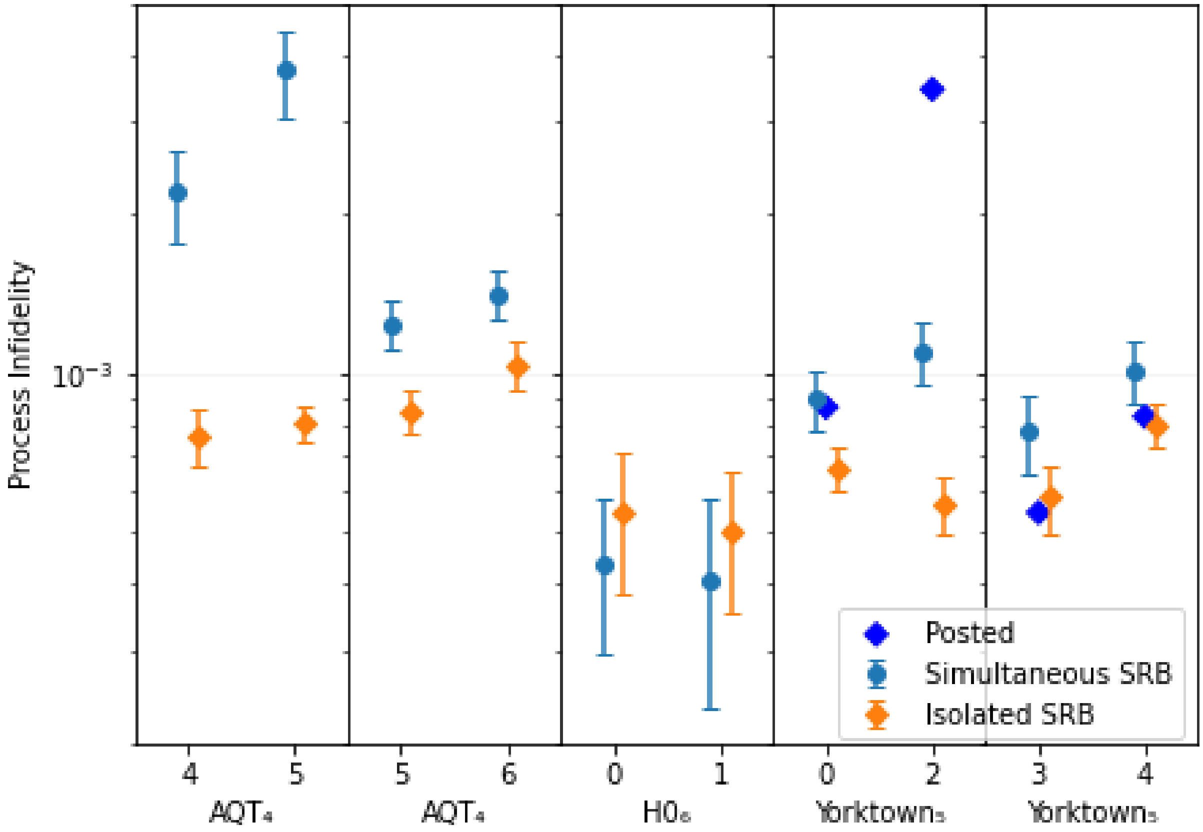


Figure 6.1: Single- qubit process infidelity of 2 Superconducting devices from *AQT_4* @ LBNL, and IBM (Yorktown$_5$), as well as an ion-trap device from Honeywell (H0$_6$), obtained by Simultaneous SRB (blue points) and Isolated SRB (orange points), including the posted estimates from IBM the morning that this data was collected (blue diamonds). Note that the x-axis is given by the qubit labels for the corresponding devices. The difference between the simultaneous and isolated SRB estimates is largely caused by the amount of error induced by the other qubit in the pair via cross-talk. As we see here, performing single-qubit gates simultaneously often reduces gate quality. Since Honeywell uses a different method to address their qubits, this analysis does not apply to their device as can be seen by the isolated values being larger than the simultaneous values. The error bars indicate a 95% confidence interval.

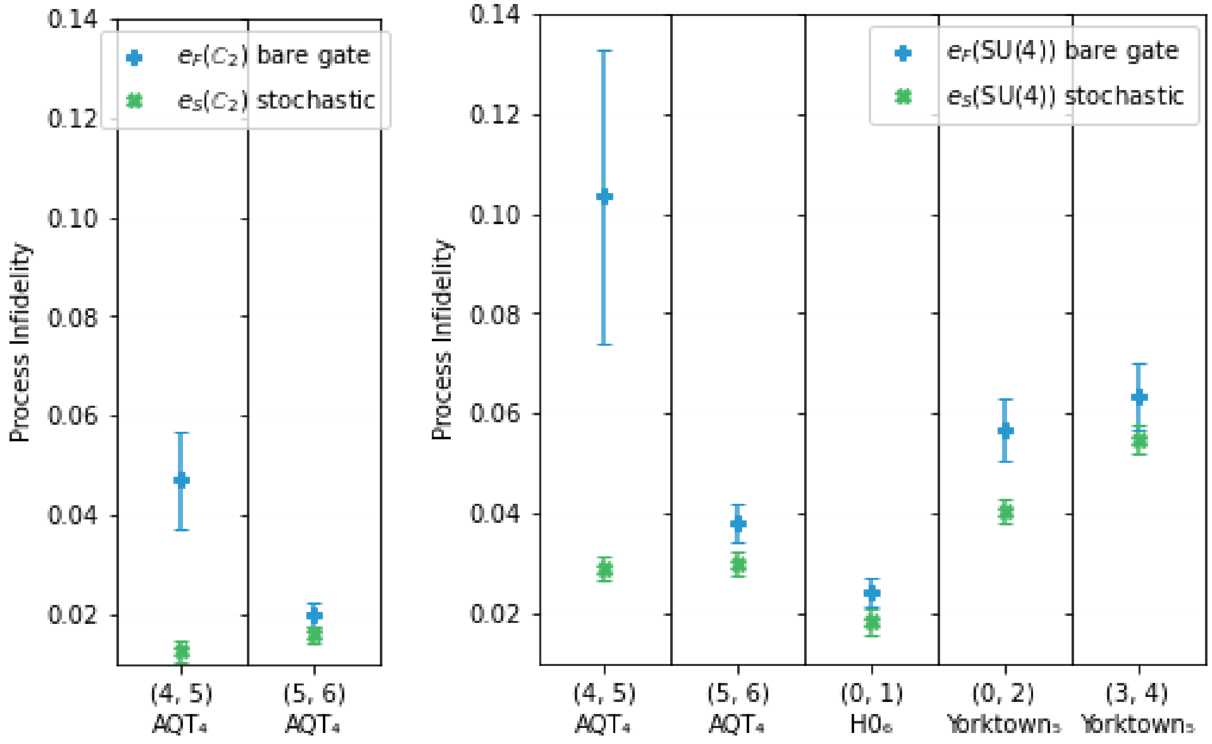
6.4.2 2-qubit results

SRB, XRB, IRB, and CB are used in conjunction to diagnose the key properties of the 2 qubit error profile, which include the bare gate $(e_F(\mathbb{C}_2), e_F(SU(4)), e_F(\mathbb{P}_2))$, stochastic $(e_S(\mathbb{C}_2), e_S(SU(4)))$, and dressed gate $(e_F(\mathbb{C}_2\mathcal{G}), e_F(SU(4)\mathcal{G}), e_F(\mathbb{P}_2\mathcal{G}))$ process infidelities, which can be used together to estimate and bound the process infidelity of the entangling gate $(e_F(\mathcal{G})_{\mathbb{C}_2}, e_F(\mathcal{G})_{SU(4)}, e_F(\mathcal{G})_{\mathbb{P}_2})$.

Estimating the stochasticity of errors by combining XRB and SRB results

As seen in fig. 6.2, the error profiles vary across platforms and even between different pairs in the same platform. For example, qubits 5 and 6 in the AQT_4 device (AQT_4(5,6)), has very low error, while the AQT_4(4,5) pair has approximately the same stochastic process infidelity, which quantifies the amount of stochastic error, as AQT_4(5,6) but is dominated by coherent error.

AQT_4(4,5) is therefore a great candidate for using Stochastic Calibration[16] to identify which unitary correction can reduce the coherent error. Stochasitc Calibration will ultimately decrease the bare gate process infidelity towards its stochastic process infidelity limit. Since AQT_4(5,6) already has a small coherent error, as quantified by the difference between the bare gate process infidelity and the stochastic process infidelity in fig. 6.2, Stochastic Calibration would not significantly improve performance for these qubits.

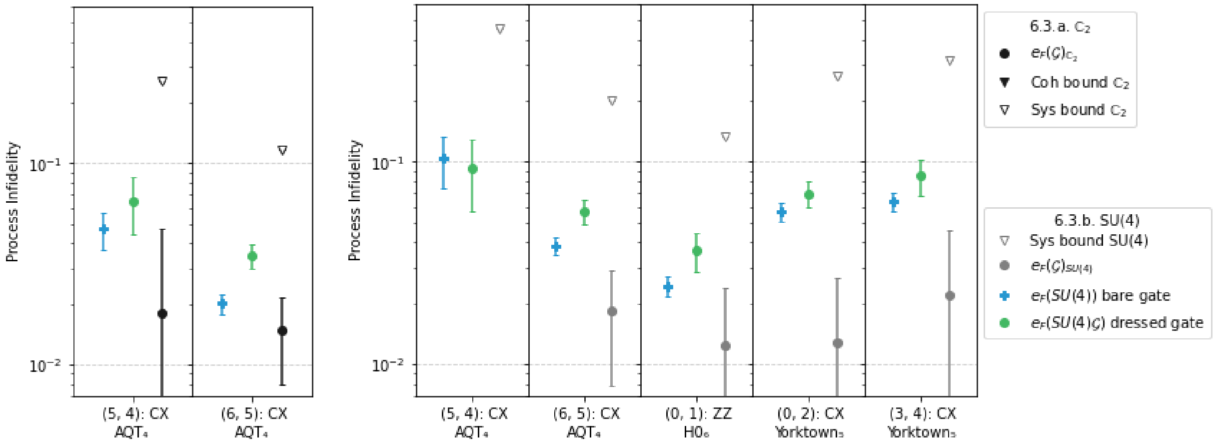


(a) Clifford Twirl

(b) SU(4) Twirl

Figure 6.2: A comparison of the bare gate process infidelity ($e_F(\mathbb{C}_2), e_F(SU(4))$) measured via SRB (blue cross), and the stochastic process infidelity ($e_S(\mathbb{C}_2), e_S(SU(4))$) measured via XRB (green X) twirling with the Clifford group (6.2a) and SU(4) group (6.2b). The difference between e_F and e_S in these figures indicates the amount of coherent error in the system(e_U). If there is a large coherent error, a significant improvement in performance can be achieved via either user-supplied correction gates (as determined via Stochastic Calibration[16]) or run-time coherent error suppression (under randomized compiling[174])

Estimating single gate infidelities by combining IRB and SRB results



(a) Clifford Twirl

(b) SU(4) Twirl

Figure 6.3: This plot demonstrates the difficulty in estimating the single gate process infidelity via SRB and IRB. The dressed gate process infidelity ($e_F(\mathbb{C}_2\mathcal{G})$ or $e_F(SU(4)\mathcal{G})$) obtained via IRB (green dot), is the average error rate of the native entangling gate composed with a random $\mathbb{C}_2$ gate (6.3a) or SU(4) (6.3b). The bare gate process infidelity ($e_F(\mathbb{C}_2)$ or $e_F(SU(4))$) from SRB (blue cross) is the average process infidelity of $\mathbb{C}_2$ (6.3a) or SU(4)(6.3b). These values can be used together to calculate (via eq. (6.7)) the single gate estimate of the native entangling gate $e_F(\mathcal{G})_{\mathbb{C}_2}$ (black dot) and $e_F(\mathcal{G})_{SU(4)}$ (grey dot) and its corresponding systematic upper bound (hollow triangle) calculated via eq. (6.8). Note that the single gate estimate is the process infidelity of the entangling gate if the coherent error of the entangling gate did not positively or negatively interfere with the coherent error of the twirling gates. The systematic bound is an estimate for if the coherent errors of the entangling gate and the twirling gates only constructively interfered (under the assumption that all of the error coherent). Here we clearly see the issue with trying to estimate the entangling gate error rate via IRB, as its value could be anything below the upper bound, which reaches as high as 0.4 for these devices.

In fig. 6.3, the single gate process infidelity estimate (shown as grey (black) points for the Clifford (SU(4)) twirling group) is the process infidelity of the entangling gate if the coherent error acting on the entangling gate did not constructively or destructively interfere with the coherent errors affecting the twirling group. The systematic upper bound in fig. 6.3 is the upper limit on the 95% confidence interval of the systematic upper bound from eq. (6.8), and represents the worst case error where the coherent error of the single gate ($\mathcal{G}$) only constructively interferes with the coherent error of the twirling gates ($\mathbb{C}_2$ and SU(4)). Therefore, this is an upper bound on the single gate process infidelity. As we will see in fig. 6.4, the coherent upper bound is tighter than the systematic upper bound because it takes into account the amount of stochastic error. This means that by not assuming all of the error is coherent (as is the case for the systematic bound), there is less coherent error that can constructively interfere, and thus the bound is tighter. It is important to tighten the upper bound such that we can learn more about the error profile of the entangling gate.

Due to the large statistical uncertainty in some implementations, such as AQT_4(5,4), the estimated bare gate process infidelity (SRB) can be larger than the dressed gate process infidelity (IRB) (see fig. 6.3). This causes the single gate estimate to be negative, and therefore the value itself is meaningless, it only indicates that the fitting or estimate of the value is imperfect.

Each element of SU(4) ($\mathbb{C}_2$) is composed of 3 (1.5) entangling gates, and 24-40 (15-25) single-qubit gates on average, depending on how the single-qubit gates are implemented. Because there are significantly more pulses in the implementation of the SU(4) (Clifford) gates than in the implementation of the entangling gate, the error arising from SU(4) (Clifford) gates exceeds that from the entangling gates by a significant margin,

$$e_F(SU(4)) \gg e_F(\mathcal{G})_{SU(4)},$$
$$e_F(\mathbb{C}_2) \gg e_F(\mathcal{G})_{\mathbb{C}_2},$$

leaving a very small difference between the bare gate and dressed gate process infidelities.

From the systematic error bounds in fig. 6.3, we see that it is unreasonable to use IRB and SRB to estimate the single gate process infidelity of the entangling gate via the systematic upper bound. We propose two alternative methods for estimating $e_F(\mathcal{G})$. The first method is to tighten the bounds by accounting for the amount of coherent error in the system (fig. 6.4), and the second is to use a weaker twirling group whose error does not drown out the error of the entangling gate(fig. 6.5).

Estimating single gate infidelities by combining IRB, XRB, and SRB results

In fig. 6.4, the systematic error bounds tighten to a large degree by accounting for the amount of coherent error in the system. From eq. (6.9), for predominantly coherent error ($u \approx 1$), there is a negligible tightening of the bound. One concrete example where this appears is in the AQT_4(5,4) pair. We can see this by comparing the systematic and coherent upper bounds. Conversely, if the error profile is dominated by stochastic noise, this bound tightens by a large amount. This happens in Yorktown$_5$(3,4), where we see the bound tightening by a factor of 3.3. This is the upper limit on improvement we've seen thus far on studied processors.

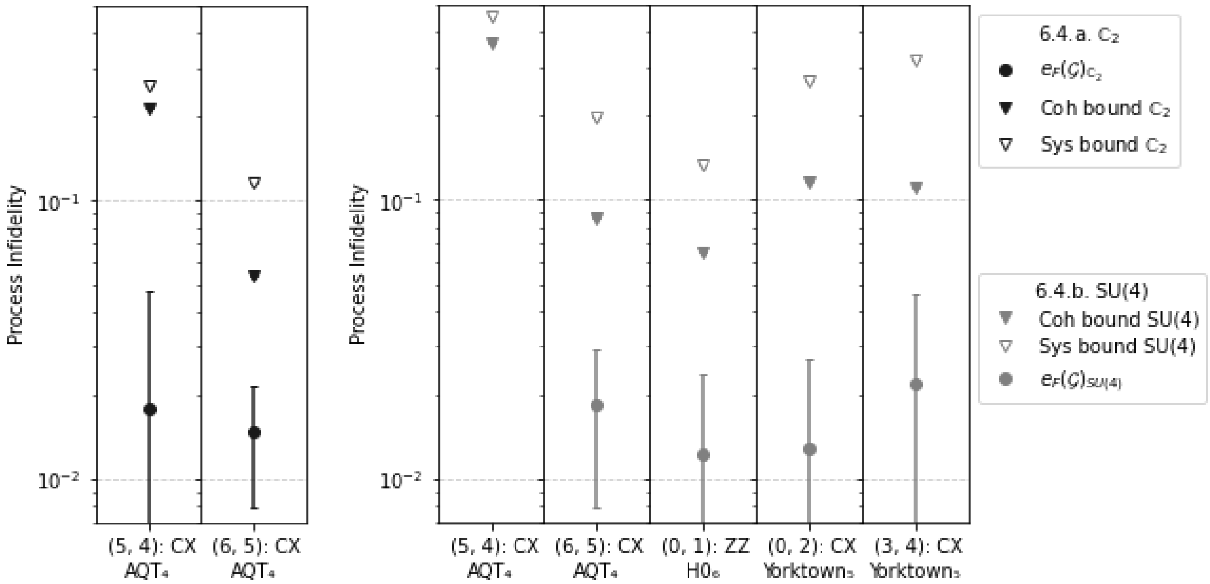
Unfortunately, even when the bound is tightened, it is not enough to allow us to make a clear estimate of the single gate process infidelity of the entangling gate due to the dominating amount of error in the twirling gates. Therefore, we propose using the Pauli group to estimate $e_F(\mathcal{G})$ to circumvent the depth issues that arise from using other twirling groups. Our results using each of these twirling groups are presented in fig. 6.5.

Estimating single gate fidelities with CB

In fig. 6.5, the systematic error bounds can be tightened considerably by using weaker randomizing sets which require fewer pulses and therefore have lower error rates. Here we see that in the experimental regime of studied processors, as the group used for the reference twirl gets weaker (SU(4)$\rightarrow \mathbb{C}_2 \rightarrow \mathbb{P}_2$), the systematic error bounds tighten and the single gate estimates have tighter statistical error bounds.

The fact that CB has tighter systematic error bounds than IRB does is even more impressive since the bounds from CB have not been tightened by taking into account the amount of coherent error in the error profile. Although not done here, it might be possible to run an XRB style experiment where $\mathbb{P}_2$ is used instead of $\mathbb{C}_2$ for the twirl. In which case the CB bounds could be tightened in the same way as the bounds from IRB have.

Even though they have been plotted together in fig. 6.5, the IRB and CB data from the Honeywell device should not be compared directly to each other as the protocols were run on different days. Specifically, they batched the set of protocols, and ran each batch on a separate day. The first batch included only the single-qubit SRB protocols from fig. 6.1, the second batch included 2 qubit SRB, XRB, and IRB, while the third batch was the two CB protocols. Note that fig. 6.5 is the only figure in which data from the same device on different days is plotted together.



(a) Clifford Twirl

(b) SU(4) Twirl

Figure 6.4: The experimental determination across various state of the art systems of the tightening of the systematic error bounds by accounting for the amount of stochastic noise in the error profile when twirling with $\mathbb{C}_2$ (6.4a) and SU(4) (6.4b) group. The upper bounds in this figure are the upper limit on the 95% confidence interval of the systematic upper bounds. The systematic upper bound calculated using IRB and SRB from eq. (6.8) (hollow black or grey triangles) are compared with the coherent upper bounds, which also take into account the unitarity from XRB from eq. (6.9) (solid black or grey triangles). This is one such attempt to solving the problem posed in fig. 6.3 to tighten the upper bound of the systematic error such that it becomes more meaningful. Here we see the bound tighten by up to a factor of 3.3 relative to the single gate infidelity estimate. Specifically, we see that when the error is mostly stochastic, the bounds tighten more than if the error is largely coherent.

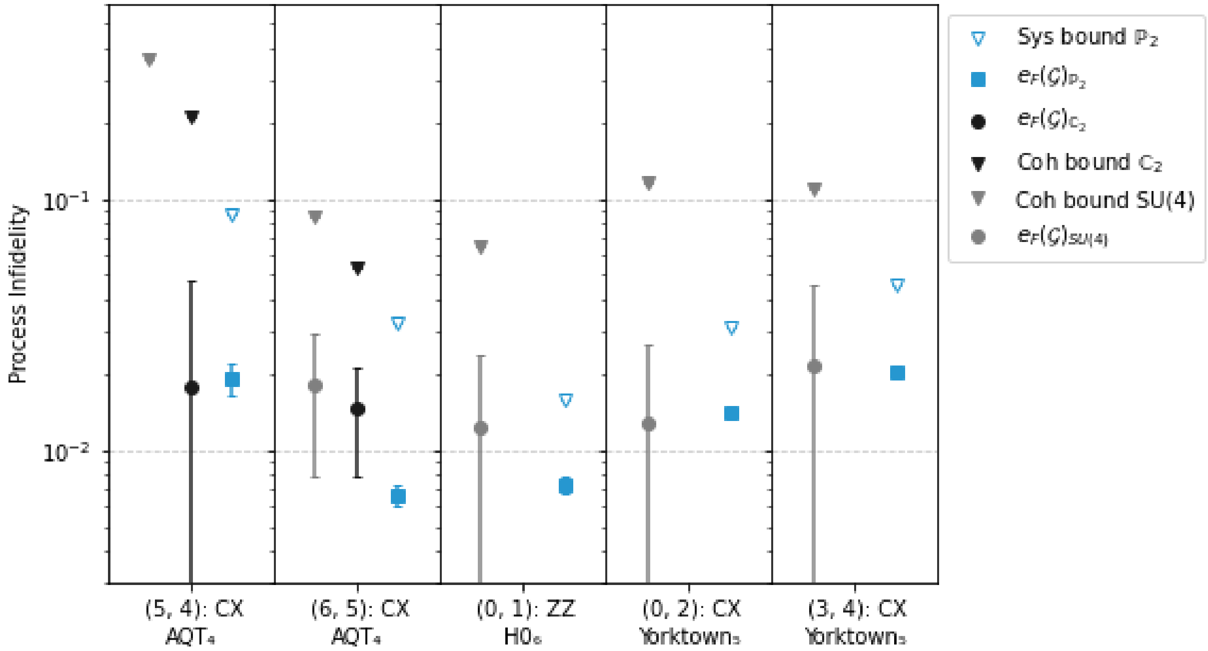


Figure 6.5: Comparison of IRB style error estimation for a gate of interest using SU(4) vs $\mathbb{C}_2$ vs $\mathbb{P}_2$, showing that the upper bound determined from combined SRB, XRB, and IRB protocols becomes increasingly tight when using randomizing sets which require fewer pulses. This implies that CB, which uses a Pauli twirl, gives the most reliable and accurate error estimates. The inferred process infidelity of the entangling gate with statistical error bounds calculated via IRB (eq. (6.7)) using SU(4) ($\mathbb{C}_2$) are shown as grey dots (black dots). The coherent upper bounds, which also take into account the unitarity from XRB (see eq. (6.9)) are depicted by the solid grey (SU(4)) triangles and solid black ($\mathbb{C}_2$) triangles. The inferred process infidelity of the entangling gate calculated via dressed gate CB and bare gate CB from eq. (6.10) are given by the blue squares, and their corresponding systematic upper bounds from eq. (6.11) are the hollow blue triangles. Here we see a significant improvement in the estimate and a dramatic tightening of the systematic upper bound by using CB instead of IRB. Clearly, even when accounting for the amount of coherent error when using IRB, CB gives a tighter bound for the estimate of $e_F(\mathcal{G})$. The error that arises when applying Pauli gates is insignificant compared to that incurred by applying the entangling gate for many of these devices, and therefore the error of the entangling gate is not washed out by the error of the Pauli twirling gates.

In fig. 6.6, we compare dressed gate process infidelity from CB to the estimate and systematic upper bound of the single gate error. Clearly, in the experimental regime where the Pauli error (bare gate CB) is insignificant compared to the error of the entangling gate, we see that the dressed gate process infidelity is a good estimate for the single gate process infidelity. Note that in the AQT_4 implementation, they physically implemented the Identity hard gates, which is why the bare gate CB values are so large. The experiments in [75] also show this behaviour where a significant portion of the total error is caused by the single qubit gates, which is not typical in other systems.

In practice the dressed entangling gate is the most relevant to accurately predicting algorithm performance, especially under Randomized Compiling, as it includes the error impact of both one round of single-qubit gates and one round of an entangling gate. We see experimentally that dressed gate CB provides the more precise statistical estimate of error, therefore dressed gate CB seems to be the best approach to a precise and accurate gold standard for characterizing gate performance.

By combining the estimates of the single-qubit gate performance, 2-qubit estimates from SRB, IRB, and XRB using a universal gate set, and the error estimate of the dressed entangling gate from dressed gate CB, we have designed a suite of benchmarking protocols that obtain the most relevant figures of merit to assess single- and 2-qubit gates.

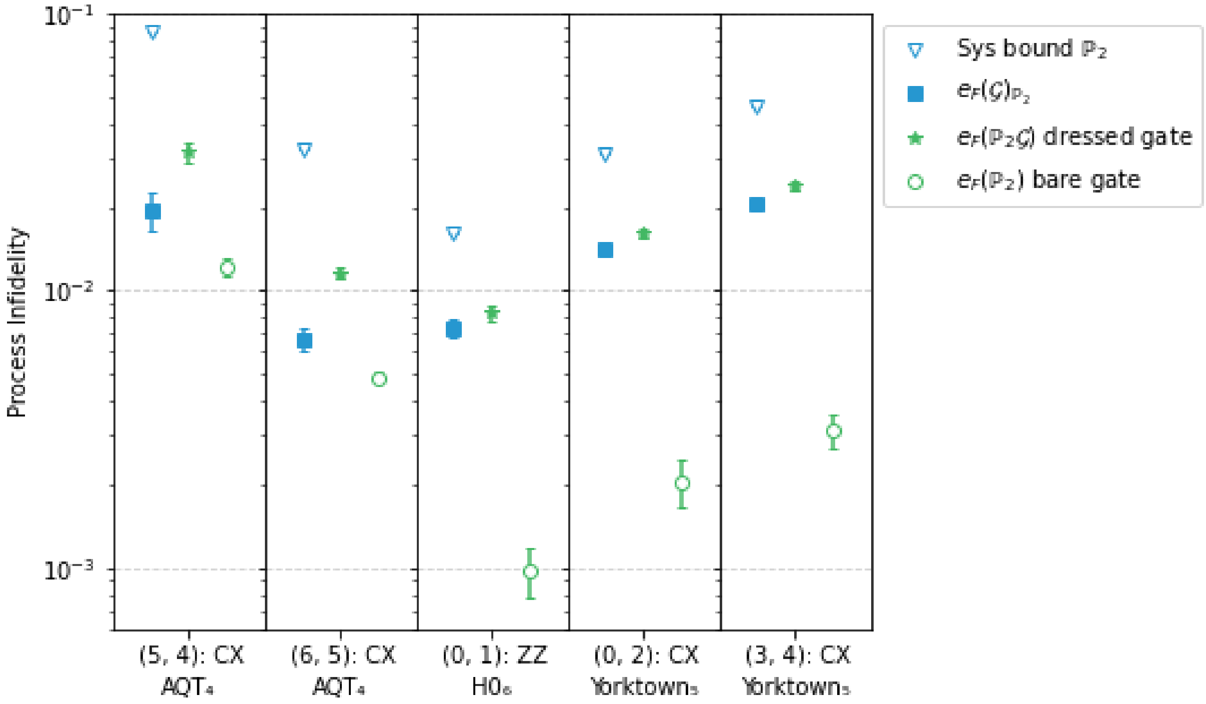


Figure 6.6: A comparison of the different approaches to estimating the inferred process infidelity of the entangling gate when using Cycle Benchmarking on various hardware platforms. The inferred process infidelity of the entangling gate calculated via dressed gate CB (green stars) and bare gate CB (green circles) from eq. (6.10) are given by the blue squares, and their corresponding systematic upper bounds from eq. (6.11) are the hollow blue triangles. In practice the dressed entangling gate is the most relevant to accurately predicting algorithm performance, especially under Randomized Compiling [174], as it includes the error impact of both one round of single-qubit gates and one round of an entangling gate and we see experimentally that dressed gate CB provides the more precise statistical estimate of error, therefore dressed gate CB is the best approach to a precise and accurate gold standard for characterizing gate performance.

6.5 Conclusion

For the first time, we have experimentally performed an unbiased, cross-platform comparison of some of the world's leading quantum computing devices. The basic gate-set assessment proposed and implemented in this chapter is a well-rounded, non-controversial standard for characterizing single- and 2-qubit gate errors. It can be used for either cross-platform comparison or local tune-up. The only difference between implementations for these two applications is which randomizing group is used. When doing an unbiased comparison, gates are randomly selected from the SU(4) group, as it is a universal gate-set and is equally difficult to implement across gate constructions. When doing local tuneup $\mathbb{C}_2$ is used due to its more standard gates. If the cross-talk is low in the system, the BGSA can be used to tune-up an entire large quantum device by testing each directly connected pair.

Although there are many user freedoms in the implementation, we gave various recommendations on how to select parameters which will optimize the accuracy of the results. The figures of merit obtained by the various protocols in the BGSA each give valuable insight into different aspects of the single- and 2-qubit error profile.

We examined the single-qubit gates of these devices and the amount of single-qubit cross-talk between the target pair of qubits. In our analysis, we observed that some devices were more impacted by cross-talk than others, and specifically that the cross-talk errors on a qubit arising from different qubits varied in magnitude even within the same device.

We also discussed the various error estimates associated with the two-qubit gates. Comparing the process infidelity of a gate-set with its stochastic error estimates, we were able to estimate the amount of coherent error affecting the device. A large coherent error indicates that it is possible to see a large improvement via user-supplied correction gates or run-time coherent error suppression.

Next, we examined the effect of interleaving the entanglement gate between the randomizing group (SU(4) or $\mathbb{C}_2$). We observed that this increases the process infidelity. The difference between the process infidelity of the bare entanglement gate and the dressed entangling gate can be used to give a rough estimate of the single gate process infidelity of the entangling gate with its corresponding systematic error bounds. In the experimental regime of studied processors, this upper bound was so large (0.1-0.4), that little could be said about the single gate process infidelity, because the error of the randomizing group (SU(4) or $\mathbb{C}_2$) drowned out the error of the entangling gate.

In the hopes of obtaining useful information about the error of the entangling gate, we tightened the bound by accounting for the amount of stochastic error in the device. We saw a noticeable improvement of the bounds for all device pairs (up to a factor of 3.3), with a larger improvement for pairs dominated by stochastic noise. If the user intends on using Randomized Compiling (RC) [174] when implementing circuits, this large upper bound would not give the most meaningful estimate of the single gate infidelity. In practise each clock cycle will be individually twirled with a Pauli gate which will cause the coherent error of the entangling gate and the twirling gates to no longer interfere.

The systematic upper bound of the entangling gate was so large because the error of the randomizing group was so large and thus had more coherent error to positively interfere with that of the entangling gate. We therefore looked at a weaker twirling group which would have significantly less error, the two qubit Pauli group, which is the basis for Cycle Benchmarking. In this regime we were able to see a significant tightening of the systematic error bound. Ultimately, the dressed gate process infidelity from Cycle Benchmarking was able to give a much more accurate estimation of the entangling gate error rate when the user intends on using RC, given that the Pauli error is insignificant in comparison to the entangling gate.

In practice, the dressed entangling gate process infidelity is the most relevant to accurately predicting algorithm performance, especially under Randomized Compiling [174, 75], as it includes the error impact of both one round of single-qubit gates and one round of an entangling gate. However, the information retrieved by the other protocols included in BGSA is also vitally important because they estimate the amount of error of universal gate-sets, the amount of stochastic error, and single-qubit gate error.

By using some of the worlds most advanced quantum computing software (TrueQ[17], Qiskit[9], and Cirq[160]), we were able to run experiments on AQT @ LBNL[121] and Honeywell [158] devices, and numerous IBM devices using IBM's remote access[159].

6.6 Supplementary Material

This section gives an overview of how to calculate the optimal number of shots that should be used for the various protocols in the BGSA, the total time (wall-clock) it takes a device to run the circuits of the BGSA, and a table which gives all of the figures of merit (and 95% confidence intervals) obtained for the 3 platforms studied in this chapter.

6.6.1 Number of shots and wall-clock time

Given the user selected parameters and the various times associated with running circuits on the specific device,

User selected parameters:

- n_{rc}: number of random circuits
- n_{shots}: number of shots per circuit
- m_j:j'th sequence length
- f_i: some decomposition/protocol-dependent fraction (average number of generators per gate usually 3 (ie 3 CNOTs to make a SWAP))

Device timing:

- t_{gi}: The average time of an i-qubit native building-block gate, including any delays between consecutive gates in a circuits (eg. x90 or CNOT or MS90)
- t_{spam}: The time (per-shot) that it takes to prepare and measure the state (and any other per-shot overhead)
- t_{load}: (t_{pick})The time it takes to load a fixed random circuits in the hardware (and any other per-fixed-circuit pulse-synthesis overhead)

we can extract the ideal number of shots needed vs the number of different circuits for a given circuit length by looking at figure 1.c. from [79]. Where:

$$t_{flip} = t_{spam} + m_j \cdot (f_1 \cdot t_{g1} + f_2 \cdot t_{g2}) \tag{6.12}$$

and $t_{load} = t_{pick}$.

Estimate	H0$_6$(0,1)	AQT$_4$(4,5)	AQT$_4$(5,6)	Yorktown$_5$(0,2)	Yorktown$_5$(3,4)
$e_F(\mathbb{C}_{q_1}\|\mathbb{C}_{q_1} \otimes \mathbb{I}_{q_2})$	5.5(1.6)e-04	7.48(0.79)e-04	8.04(0.89)e-04	6.64(0.68)e-04	5.63(0.79)e-04
$e_F(\mathbb{C}_{q_2}\|\mathbb{I}_{q_1} \otimes \mathbb{C}_{q_2})$	5.0(1.5)e-04	7.70(0.77)e-04	1.08(0.13)e-03	6.25(0.77)e-04	7.73(0.82)e-04
$e_F(\mathbb{C}_{q_1}\|\mathbb{C}_{q_1} \otimes \mathbb{C}_{q_2})$	4.3(1.4)e-04	2.22(0.41)e-03	1.31(0.18)e-03	8.8(1.2)e-04	7.8(1.3)e-04
$e_F(\mathbb{C}_{q_2}\|\mathbb{C}_{q_1} \otimes \mathbb{C}_{q_2})$	4.1(1.7)e-04	3.81(0.71)e-03	1.46(0.18)e-03	9.8(1.4)e-04	1.15(0.15)e-03
$e_F(SU(4))$	2.43(0.27)e-02	1.04(0.29)e-01	3.83(0.38)e-02	5.68(0.63)e-02	6.35(0.69)e-02
$e_F(\mathbb{C}_2)$		4.71(0.98)e-02	2.01(0.23)e-02		
$e_S(SU(4))$	1.83(0.26)e-02	2.91(0.25)e-02	3.00(0.24)e-02	4.04(0.24)e-02	5.50(0.3)e-02
$e_S(\mathbb{C}_2)$		1.25(0.2)e-02	1.60(0.16)e-02		
$e_F(SU(4)\mathcal{G})$	3.66(0.81)e-02	9.3(3.6)e-02	5.68(0.75)e-02	6.96(0.98)e-02	8.5(1.7)e-02
$e_F(\mathbb{C}_2\mathcal{G})$		6.5(2.1)e-02	3.48(0.48)e-02		
$e_F(\mathcal{G})_{SU(4)}$	1.2(1.1)e-02		1.8(1.1)e-02	1.3(1.4)e-02	2.2(2.4)e-02
$e_F(\mathcal{G})_{\mathbb{C}_2}$		1.8(2.9)e-02	1.47(0.68)e-02		
Sys Bound SU(4)	1.19(0.15)e-01	3.72(0.83)e-01	1.84(0.15)e-01	2.44(0.21)e-01	2.84(0.32)e-01
Coh Bound SU(4)	4.6(1.9)e-02	2.73(0.9)e-01	6.6(1.9)e-02	9.2(2.5)e-02	7.4(3.7)e-02
Sys Bound $\mathbb{C}_2$		2.17(0.42)e-01	1.06(0.1)e-01		
Coh Bound $\mathbb{C}_2$		1.71(0.43)e-01	4.1(1.3)e-02		
$e_F(\mathbb{P}_2\mathcal{G})$	8.30(0.53)e-03	3.16(0.28)e-02	1.156(0.06)e-02	1.615(0.064)e-02	2.377(0.092)e-02
$e_F(\mathbb{P}_2)$	9.8(2.0)e-04	1.22(0.1)e-02	4.86(0.25)e-03	2.05(0.4)e-03	3.13(0.43)e-03
$e_F(\mathcal{G})_{\mathbb{P}_2}$	7.32(0.56)e-03	1.95(0.29)e-02	6.69(0.65)e-03	1.410(0.076)e-02	2.06(0.1)e-02
Sys Bound $\mathbb{P}_2$	1.49(0.1)e-02	8.15(0.5)e-02	3.12(0.12)e-02	2.95(0.17)e-02	4.38(0.2)e-02

Table 6.4: All of the estimates and 95% confidence intervals from the BGSA protocols. Note that "Sys Bound" stands for systematic upper bound from eq. (6.8), while "Coh Bound" stand for Coherent upper bound from eq. (6.9). Note notation $e_F(\mathcal{G})_{\mathbb{G}}$, where this is the estimate of $e_F(\mathcal{G})$, given that the randomizing set $\mathbb{G}$ was used.

Once the optimal number of shots has been estimated ($N_{optimal}$ from [79]), we can then approximate the wall-clock time of performing an RB+ experiment as:

$$T = \sum_{j} n_{rc} \cdot (t_{load} + n_{shots} \cdot (t_{spam} + m_j \cdot (f_1 \cdot t_{g1} + f_2 \cdot t_{g2}))) \tag{6.13}$$

6.6.2 Acknowledgements

We acknowledge use of the IBM Q for this work. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Q team.

Chapter 7

Conclusion

In this thesis, we began in chapter 3 by looking at the physical interpretation of the unitarity from XRB under physically realistic gate-dependent errors. We showed both analytically and numerically that the unitarity obtained by running XRB on a gate-set with gate-dependent noise was the average unitarity of each gate and not the unitarity of the average gate. This is significant because the average unitarity of each gate is able to differentiate between gate-dependent coherent errors (e.g., due to miscalibration) and incoherent errors (e.g., intrinsic noise processes). Since the unitarity was first proposed as an estimate for the proportion of noise which is coherent, this work verifies that when the gate-independent error assumption is lifted, this figure of merit retains its value.

In chapter 4, we turned our sights to the physical interpretation of the decay constant (p) from SRB. We showed that in a physically relevant limit, p can be used to calculate a novel figure of merit, the gate-set circuit fidelity. As defined in eq. (4.4), p represents the relative change in the gate-set circuit fidelity when an additional random noisy operation from $\tilde{\mathbb{G}}$ is introduced to a random circuit constructed from elements in $\tilde{\mathbb{G}}$ eq. (4.12). We demonstrated analytically that p is – up to the second order in the infidelity – in one-to-one correspondence with a physical gate-set fidelity. We further showed constructively which physical basis change on a given reference gate-set would be needed to obtain the reconciliation between p and the average gate-set fidelity as in eq. (4.13). Finally, we provided numerical evidence and physically motivated arguments to support the conjecture that our results extend to multiple qubits.

The importance of randomized benchmarking (RB) as a tool for estimating error rates in elementary gate operations is devalued by the conflation of SRB and NIST RB protocols in the literature, which can cause problems due to the distinct outcomes generated by the respective protocols. The rigorous theoretical framework developed in chapter 5 proves that NIST RB, like SRB, leads to an exponential decay which depends only on the underlying gate-independent error model. We have also demonstrated that SRB and NIST RB, which sample from a uniform 2-design and a subset respectively, lead to significantly different observed error rates in many physically realistic gate-dependent error models and pulse sets. This is because the figures of merit that these protocols evaluate have physically different meanings; for SRB, its the infidelity of $\mathbb{C}$ and for NIST RB it is the infidelity of a $\pi/2$ pulse followed by a $\mathbb{P}$. These rates can differ threefold, and can therefore yield significantly different recommendations regarding the amount of overhead burden in the implementation of fault-tolerant quantum error correction. From this project it became clear that the only way to compare different devices was to set a standard for how to diagnose errors in basic gate-sets, by clearly defining input parameters, gate-sets, and protocols. This led us to develop the basic gate-set assessment, covered in chapter 6.

In chapter 6, we propose a characterization routine which we refer to as the basic gate-set assessment (BGSA). The BGSA consists of a suite of benchmarking protocols together with data interpretation guidelines for which the goal is to give a thorough assessment of the single- and two-qubit error profile of a universal gate-set. Using the BGSA, we implemented the first even-handed, cross-platform comparison of some of the world's leading quantum computing devices. Using some of the world's most advanced quantum computing software (TrueQ, qiskit, and Cirq), we ran this suite of protocols on devices from several leading developers of quantum computing hardware, including the AQT@LBNL and Honeywell devices, and numerous IBM devices using IBM's remote access. In addition to the cross-platform comparison application, the BGSA can also be used for local tune-up by using a randomizing group which is not a universal gate-set and whose elements can be implemented using fewer control pulses. If the cross-talk error rate is sufficiently low, the BGSA can be used to optimize performance even in large devices. In order to avoid confusion about the implementation details of the BGSA, as we saw with RB in chapter 5, we discuss in detail the many user freedoms in the implementation of the BGSA and provide recommendations for each parameter under each use-case.

In our analysis of the single-qubit gate error, we quantified the severity of single-qubit cross-talk between different target pairs of qubits. We observed that the significance of cross-talk errors varied for different devices and qubit pairs. We compared the various error estimates associated with the two-qubit gate error. This analysis includes a comparison of the process infidelity of the SU(4) and $\mathbb{C}_2$ gate-sets with the corresponding stochastic error estimates, in order to estimate the amount of coherent error affecting each device. From these estimates, we could quantify the extent to which a device could benefit from user-supplied correction gates or run-time coherent error suppression.

We compared bare and dressed gate process infidelities from SRB and IRB to upper bound the entangling gate process infidelity. In the experimental regime of studied processors, the error of SU(4) and $\mathbb{C}_2$ gates far outweighed that of the entangling gate, and so the infidelity upper bound was consistently too large for anything to be said about the error of the entangling gate. We therefore experimentally implemented for the first time the method first proposed in [34] to tighten the infidelity upper bound by accounting for the amount of systematic error. Despite a significant decrease of the upper bound, our improved figure was still not sufficiently low to provide meaningful information about single gate process infidelity. Ultimately, we were able to use a randomizing set requiring fewer pulses ($\mathbb{P}_2$) to obtain meaningful upper bound on the entangling gate process infidelity. We then compared these estimates to the dressed gate process infidelity from CB and noted that if the error of the $\mathbb{P}_2$ gates is sufficiently low, this should be used as an estimate of the process infidelity of the entangling gate.

References

[1] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007.

[2] Deanna M Abrams, Nicolas Didier, Shane A Caldwell, Blake R Johnson, and Colm A Ryan. Methods for measuring magnetic flux crosstalk between tunable transmons. *Physical Review Applied*, 12(6):064022, 2019.

[3] Dorit Aharonov and Michael Ben-Or. Fault-Tolerant Quantum Computation With Constant Error Rate. *SIAM J. Comput.*, 38(4):1207, 1999.

[4] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error rate. *SIAM Journal on Computing*, 2008.

[5] Joseph B Altepeter, David Branning, Evan Jeffrey, TC Wei, Paul G Kwiat, Robert T Thew, Jeremy L OBrien, Michael A Nielsen, and Andrew G White. Ancilla-assisted quantum process tomography. *Physical Review Letters*, 90(19):193601, 2003.

[6] 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.

[7] 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.

[8] 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. Supplementary information for" quantum supremacy using a programmable superconducting processor". *arXiv preprint arXiv:1910.11333*, 2019.

[9] Abraham Asfaw, Luciano Bello, Yael Ben-Haim, Sergey Bravyi, Nicholas Bronn, Lauren Capelluto, Almudena Carrera Vazquez, Jack Ceroni, Richard Chen, Albert Frisch, Jay Gambetta, Shelly Garion, Leron Gil, Salvador De La Puente Gonzalez, Francis Harkins, Takashi Imamichi, David McKay, Antonio Mezzacapo, Zlatko Minev, Ramis Movassagh, Giacomo Nannicni, Paul Nation, Anna Phan, Marco Pistoia, Arthur Rattew, Joachim Schaefer, Javad Shabani, John Smolin, John Stenger, Kristan Temme, Madeleine Tod, Stephen Wood, and James Wootton. Learn quantum computation using qiskit, 2020.

[10] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O/'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and John M. Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500–503, 2014. Letter.

[11] R Barends, J Kelly, A Megrant, A Veitia, D Sank, E Jeffrey, TC White, J Mutus, AG Fowler, B Campbell, et al. Logic gates at the surface code threshold: Superconducting qubits poised for fault-tolerant quantum computing. *arXiv preprint arXiv:1402.4848*, 2014.

[12] R Barends, J Kelly, A Veitia, A Megrant, AG Fowler, B Campbell, Y Chen, Z Chen, B Chiaro, A Dunsworth, et al. Rolling quantum dice with a superconducting qubit. *Physical Review A*, 90(3):030303, 2014.

[13] Panagiotis Kl Barkoutsos, Fotios Gkritsis, Pauline J Ollitrault, Igor O Sokolov, Stefan Woerner, and Ivano Tavernelli. Quantum algorithm for alchemical optimization in material design. *arXiv preprint arXiv:2008.06449*, 2020.

[14] F. L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2(1):137, 1960.

[15] Stefanie J. Beale, Arnaud Carignan-Dugas, Dar Dahlen, Joseph Emerson, Ian Hincks, Pavithran Iyer, Aditya Jain, David Hufnagel, Egor Ospadov, Jordan Saunders, Andrew Stasiuk, Joel J. Wallman, and Adam Winick. True-q rcal. `https://trueq.quantumbenchmark.com/api/protocols.html?highlight=rcal#trueq.make_rcal`.

[16] Stefanie J. Beale, Arnaud Carignan-Dugas, Dar Dahlen, Joseph Emerson, Ian Hincks, Pavithran Iyer, Aditya Jain, David Hufnagel, Egor Ospadov, Jordan Saunders, Andrew Stasiuk, Joel J. Wallman, and Adam Winick. True-q stochas-

tic calibration. https://trueq.quantumbenchmark.com/guides/protocols/sc.html?highlight=make_sc.

[17] Stefanie J. Beale, Arnaud Carignan-Dugas, Dar Dahlen, Joseph Emerson, Ian Hincks, Pavithran Iyer, Aditya Jain, David Hufnagel, Egor Ospadov, Jordan Saunders, Andrew Stasiuk, Joel J. Wallman, and Adam Winick. True-q, 2020.

[18] Bruno Bellomo, Antonella De Pasquale, Giulia Gualdi, and Ugo Marzolino. Reconstruction of markovian master equation parameters through symplectic tomography. *Physical Review A*, 80:052108, 2009.

[19] Lev S Bishop, Sergey Bravyi, Andrew Cross, Jay M Gambetta, and John Smolin. Quantum volume. *Quantum Volume. Technical Report*, 2017.

[20] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Jonathan Mizrahi, Jonathan D. Sterk, and Peter Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, 2013.

[21] Robin Blume-Kohout, Kenneth Rudinger, Erik Nielsen, Timothy Proctor, and Kevin Young. Wildcard error: Quantifying unmodeled errors in quantum processors. *Bulletin of the American Physical Society*, 2021.

[22] Robin Blume-Kohout and Kevin C Young. A volumetric framework for quantum computer benchmarks. *Quantum*, 4:362, 2020.

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

[24] Max Born. Quantenmechanik der stoßvorgänge. *Zeitschrift für Physik*, 38(11-12):803–827, 1926.

[25] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum*, 1:8, 2017.

[26] Kenton R Brown, Andrew C Wilson, Yves Colombe, C Ospelkaus, Adam M Meier, E Knill, D Leibfried, and David J Wineland. Single-qubit-gate error below 10- 4 in a trapped ion. *Physical Review A*, 84(3):030303, 2011.

[27] Winton G Brown and Bryan Eastin. Randomized benchmarking with restricted gate sets. *Physical Review A*, 97(6):062323, 2018.

[28] Donovan Buterakos, Robert E Throckmorton, and S Das Sarma. Crosstalk error correction through dynamical decoupling of single-qubit gates in capacitively coupled singlet-triplet semiconductor spin qubits. *Physical Review B*, 97(4):045431, 2018.

[29] Yudong Cao, Jhonathan Romero, and Alán Aspuru-Guzik. Potential of quantum computing for drug discovery. *IBM Journal of Research and Development*, 62(6):6–1, 2018.

[30] Arnaud Carignan-Dugas, Matthew Alexander, and Joseph Emerson. A polar decomposition for quantum channels (with applications to bounding error propagation in quantum circuits). *Quantum*, 3:173, 2019.

[31] Arnaud Carignan-Dugas, Kristine Boone, Joel J Wallman, and Joseph Emerson. From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters. *New Journal of Physics*, 2018.

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

[33] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. Efficiently characterizing the total error in quantum circuits. *arXiv preprint arXiv:1610.05296*, 36, 2016.

[34] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. Bounding the average gate fidelity of composite channels using the unitarity. *New Journal of Physics*, 21(5):053016, 2019.

[35] L. Casparis, T. W. Larsen, M. S. Olsen, F. Kuemmeth, P. Krogstrup, J. Nygârd, K. D. Petersson, and C. M. Marcus. Gatemon Benchmarking and Two-Qubit Operations. *Physical Review Letters*, 116(15):150505, 2016.

[36] Pascal Cerfontaine, Ren Otten, and Hendrik Bluhm. Self-consistent calibration of quantum-gate sets. *Physical Review Applied*, 13(4), 2020.

[37] Tobias Chasseur and Frank K Wilhelm. Complete randomized benchmarking protocol accounting for leakage errors. *Physical Review A*, 92(4):042333, 2015.

[38] Kevin S Chou, Jacob Z Blumoff, Christopher S Wang, Philip C Reinhold, Christopher J Axline, Yvonne Y Gao, Luigi Frunzio, MH Devoret, Liang Jiang, and RJ Schoelkopf. Deterministic teleportation of a quantum gate between two logical qubits. *Nature*, 561(7723):368–373, 2018.

[39] JM Chow, Jay M Gambetta, Lars Tornberg, Jens Koch, Lev S Bishop, Andrew A Houck, BR Johnson, L Frunzio, Steven M Girvin, and Robert J Schoelkopf. Randomized benchmarking and process tomography for gate errors in a solid-state qubit. *Physical review letters*, 102(9):090502, 2009.

[40] Isaac L Chuang and Michael A Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997.

[41] Isaac L. Chuang and Michael A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455, 1997.

[42] Antonio D. Córcoles, Jay M. Gambetta, Jerry M. Chow, John A. Smolin, Matthew Ware, Joel Strand, B. L. T. Plourde, and Matthias Steffen. Process verification of two-qubit quantum gates by randomized benchmarking. *Physical Review A*, 87(3):030301, 2013.

[43] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta. Scalable randomised benchmarking of non-Clifford gates. *npj Quantum Information*, 2:16012, 2016.

[44] Andrew W Cross, Lev S Bishop, Sarah Sheldon, Paul D Nation, and Jay M Gambetta. Validating quantum computers using randomized model circuits. *Physical Review A*, 100(3):032328, 2019.

[45] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs: Constructions and applications. *arXiv preprint quant-ph/0606161*, 2006.

[46] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009.

[47] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.

[48] Leonardo DiCarlo, Jerry M Chow, Jay M Gambetta, Lev S Bishop, Blake R Johnson, DI Schuster, J Majer, Alexandre Blais, Luigi Frunzio, SM Girvin, et al. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature*, 460(7252):240–244, 2009.

[49] 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.

[50] Joseph Emerson, Marcus Silva, Osama Moussa, Colm A. Ryan, Martin Laforest, Jonathan Baugh, David G Cory, and Raymond Laflamme. Symmetrized characterization of noisy quantum processes. *Science*, 317(5846):1893–6, 2007.

[51] Suguru Endo, Simon C Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, 8(3):031027, 2018.

[52] Alexander Erhard, Joel James Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban Adrian Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. Characterizing large-scale quantum computers via cycle benchmarking. *arXiv preprint arXiv:1902.08543*, 2019.

[53] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm. *arXiv preprint arXiv:1602.07674*, 2016.

[54] Aleksey K Fedorov, Evgeniy O Kiktenko, and Alexander I Lvovsky. Quantum computers put blockchain security at risk, 2018.

[55] Guanru Feng, Joel J Wallman, Brandon Buonacorsi, Franklin H Cho, Daniel K Park, Tao Xin, Dawei Lu, Jonathan Baugh, and Raymond Laflamme. Estimating the coherence of noise in quantum control of a solid-state qubit. *Physical Review Letters*, 117(26):260501, 2016.

[56] Christopher Ferrie. Self-guided quantum tomography. *Physical review letters*, 113(19):190404, 2014.

[57] Caroline Figgatt, Dmitri Maslov, Kevin A Landsman, Norbert Matthias Linke, Shantanu Debnath, and Christofer Monroe. Complete 3-qubit grover search on a programmable quantum computer. *Nature Communications*, 8(1):1–9, 2017.

[58] Jaromír Fiurášek and Zdeněk Hradil. Maximum-likelihood estimation of quantum processes. *Physical Review A*, 63(2):020101, 2001.

[59] Steven T Flammia and Joel J Wallman. Efficient estimation of pauli channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, 2020.

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

[61] Christopher A Fuchs and Jeroen Van De Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.

[62] 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.

[63] W Gale, E Guth, and GT Trammell. Determination of the quantum state by measurements. *Physical Review*, 165(5):1434, 1968.

[64] Jay M Gambetta, Antonio D Córcoles, Seth T Merkel, Blake R Johnson, John A Smolin, Jerry M Chow, Colm A Ryan, Chad Rigetti, S Poletto, Thomas A Ohki, et al. Characterization of addressability by simultaneous randomized benchmarking. *Physical Review Letters*, 109(24):240504, 2012.

[65] Xun Gao, Sheng-Tao Wang, and L-M Duan. Quantum supremacy for simulating a translation-invariant ising spin model. *Physical Review Letters*, 118(4):040502, 2017.

[66] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[67] Lov K Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical review letters*, 79(2):325, 1997.

[68] Erwin L Hahn. Spin echoes. *Physical review*, 80(4):580, 1950.

[69] Brian C Hall. *Quantum theory for mathematicians.* Springer, 2013.

[70] D Hangleiter, M Kliesch, M Schwarz, and J Eisert. Direct certification of a class of quantum simulations. *Quantum Science and Technology*, 2(1):015004, 2017.

[71] Robin Harper, Steven T Flammia, and Joel J Wallman. Efficient learning of quantum noise. *Nature Physics*, 16(12):1184–1188, 2020.

[72] Robin Harper, Wenjun Yu, and Steven T Flammia. Fast estimation of sparse quantum noise. *PRX Quantum*, 2(1):010322, 2021.

[73] TP Harty, DTC Allcock, C J Ballance, L Guidoni, HA Janacek, NM Linke, DN Stacey, and DM Lucas. High-fidelity preparation, gates, memory, and readout of a trapped-ion quantum bit. *Physical Review Letters*, 113(22):220501, 2014.

[74] AK Hashagen, ST Flammia, D Gross, and JJ Wallman. Real randomized benchmarking. *Quantum*, 2:85, 2018.

[75] Akel Hashim, Ravi K. Naik, Alexis Morvan, Jean-Loup Ville, Bradley Mitchell, John Mark Kreikebaum, Marc Davis, Ethan Smith, Costin Iancu, Kevin P. O'Brien, Ian Hincks, Joel J. Wallman, Joseph Emerson, and Irfan Siddiqi. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, 2020.

[76] Jonas Helsen, Ingo Roth, Emilio Onorati, Albert H. Werner, and Jens Eisert. A general framework for randomized benchmarking. 2020.

[77] Jonas Helsen, Joel J Wallman, Steven T Flammia, and Stephanie Wehner. Multiqubit randomized benchmarking using few samples. *Physical Review A*, 100(3):032304, 2019.

[78] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P Lanyon, Peter Love, Ryan Babbush, et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X*, 8(3):031022, 2018.

[79] Ian Hincks, Joel J Wallman, Chris Ferrie, Chris Granade, and David G Cory. Bayesian inference for randomized benchmarking protocols. *arXiv preprint arXiv:1802.00401*, 2018.

[80] Holger F. Hofmann. Complementary classical fidelities as an efficient criterion for the evaluation of experimentally realized quantum operations. *Physical Review Letters*, 94:160504, 2005.

[81] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge University Press, 2012.

[82] Zdenek Hradil. Quantum-state estimation. *Physical Review A*, 55(3):R1561, 1997.

[83] W Huang, CH Yang, KW Chan, T Tanttu, B Hensen, RCC Leon, MA Fogarty, JCC Hwang, FE Hudson, Kohei M Itoh, A Morello, A Laucht, and A. S. Dzurak. Fidelity benchmarks for two-qubit gates in silicon. *Nature*, 569(7757):532–536, 2019.

[84] Daniel FV James, Paul G Kwiat, William J Munro, and Andrew G White. On the measurement of qubits. In *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, pages 509–538. World Scientific, 2005.

[85] Julian Kelly, R Barends, B Campbell, Y Chen, Z Chen, B Chiaro, A Dunsworth, Austin G Fowler, I-C Hoi, E Jeffrey, et al. Optimal quantum control using randomized benchmarking. *Physical Review Letters*, 112(24):240504, 2014.

[86] Julian Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, Austin G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, Chris Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis. Optimal Quantum Control Using Randomized Benchmarking. *Physical Review Letters*, 112(24):240504, 2014.

[87] Dohun Kim, Zhan Shi, CB Simmons, DR Ward, JR Prance, Teck Seng Koh, John King Gamble, DE Savage, MG Lagally, Mark Friesen, et al. Quantum control and process tomography of a semiconductor quantum dot hybrid qubit. *Nature*, 511(7507):70–74, 2014.

[88] Shelby Kimmel, Marcus P da Silva, Colm A Ryan, Blake R Johnson, and Thomas Ohki. Robust extraction of tomographic information via randomized benchmarking. *Physical Review X*, 4(1):011050, 2014.

[89] Shelby Kimmel, Marcus P. da Silva, Colm a. Ryan, Blake R. Johnson, and Thomas a. Ohki. Robust Extraction of Tomographic Information via Randomized Benchmarking. *Physical Review X*, 4(1):011050, 2014.

[90] Shelby Kimmel, Guang Hao Low, and Theodore J Yoder. Robust calibration of a universal single-qubit gate set via robust phase estimation. *Physical Review A*, 92(6):062315, 2015.

[91] A Yu Kitaev. Quantum error correction with imperfect gates. In O. Hirota, A.S. Holevo, and C.M. Caves, editors, *Quantum Communication, Computing, and Measurement*, pages 181–188. Springer, 1997.

[92] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, 2008.

[93] Emanuel Knill. Resilient Quantum Computation. *Science*, 279(5349):342–345, 1998.

[94] Emanuel Knill. Quantum computing with realistically noisy devices. *Nature*, 434(7029):39, 2005.

[95] Emanuel Knill, D Leibfried, R Reichle, J Britton, RB Blakestad, John D Jost, C Langer, R Ozeri, Signe Seidelin, and David J Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, 2008.

[96] Ben P Lanyon, Till J Weinhold, Nathan K Langford, Marco Barbieri, Daniel FV James, Alexei Gilchrist, and Andrew G White. Experimental demonstration of a compiled version of Shor's algorithm with quantum entanglement. *Physical Review Letters*, 99(25):250505, 2007.

[97] BP Lanyon, C Maier, Milan Holzäpfel, Tillmann Baumgratz, C Hempel, P Jurcevic, Ish Dhand, AS Buyskikh, AJ Daley, Marcus Cramer, et al. Efficient tomography of a quantum many-body system. *Nature Physics*, 13(12):1158–1162, 2017.

[98] Ludovico Latmiral, Nicolò Spagnolo, and Fabio Sciarrino. Towards quantum supremacy with lossy scattershot boson sampling. *New Journal of Physics*, 18(11):113008, 2016.

[99] Benjamin Levi, Cecilia C Lopez, Joseph Emerson, and David G. Cory. Efficient error characterization in quantum information processing. *Physical Review A*, 75:022314, 2007.

[100] Junan Lin, Brandon Buonacorsi, Raymond Laflamme, and Joel J Wallman. On the freedom in representing quantum operations. *New Journal of Physics*, 21(2):023006, 2019.

[101] Norbert M Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A Landsman, Kenneth Wright, and Christopher Monroe. Experimental comparison of two quantum computing architectures. *Proceedings of the National Academy of Sciences*, 114(13):3305–3310, 2017.

[102] Mirko Lobino, Dmitry Korystov, Connor Kupchak, Eden Figueroa, Barry C Sanders, and AI Lvovsky. Complete characterization of quantum-optical processes. *Science*, 322(5901):563–566, 2008.

[103] Austin P Lund, Michael J Bremner, and Timothy C Ralph. Quantum sampling problems, bosonsampling and quantum supremacy. *npj Quantum Information*, 3(1):1–8, 2017.

[104] JS Lundeen, A Feito, H Coldenstrodt-Ronge, KL Pregnell, Ch Silberhorn, TC Ralph, J Eisert, MB Plenio, and IA Walmsley. Tomography of quantum detectors. *Nature Physics*, 5(1):27–30, 2009.

[105] Ruichao Ma, Brendan Saxberg, Clai Owens, Nelson Leung, Yao Lu, Jonathan Simon, and David I Schuster. A dissipatively stabilized mott insulator of photons. *Nature*, 566(7742):51–57, 2019.

[106] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Scalable and Robust Randomized Benchmarking of Quantum Processes. *Physical Review Letters*, 106(18):180504, 2011.

[107] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Physical Review A*, 85(4):042311, 2012.

[108] Easwar Magesan, Jay M. Gambetta, Blake R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, Mark B. Ketchen, and Matthias Steffen. Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking. *Physical Review Letters*, 109(8):080505, 2012.

[109] S Mavadia, CL Edmunds, C Hempel, H Ball, F Roy, TM Stace, and MJ Biercuk. Experimental quantum verification in the presence of temporally correlated noise. *NPJ Quantum information*, 4(1):1–9, 2018.

[110] Alexander J McCaskey, Zachary P Parks, Jacek Jakowski, Shirley V Moore, Titus D Morris, Travis S Humble, and Raphael C Pooser. Quantum chemistry as a benchmark for near-term quantum computers. *npj Quantum Information*, 5(1):1–8, 2019.

[111] D. C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta. Universal Gate for Fixed-Frequency Qubits via a Tunable Bus. *Physical Review Applied*, 6(6):064007, 2016.

[112] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta. Three Qubit Randomized Benchmarking. *ArXiv e-prints*, 2017.

[113] David C McKay, Sarah Sheldon, John A Smolin, Jerry M Chow, and Jay M Gambetta. Three qubit randomized benchmarking. *arXiv preprint arXiv:1712.06550*, 2017.

[114] A. M. Meier. *Randomized Benchmarking of Clifford Operators*. PhD thesis, University of Colorado, Boulder, 2013.

[115] Seth T. Merkel, Jay M. Gambetta, John A. Smolin, Stefano Poletto, Antonio D. Córcoles, Blake R. Johnson, Colm A. Ryan, and Matthias Steffen. Self-consistent quantum process tomography. *Physical Review A*, 87:062119, 2013.

[116] Seth T Merkel, Emily J Pritchett, and Bryan H Fong. Randomized benchmarking as convolution: Fourier analysis of gate dependent errors. *arXiv preprint arXiv:1804.05951*, 2018.

[117] M. Mohseni and D. A. Lidar. Direct characterization of quantum dynamics. *Physical Review Letters*, 97:170501, 2006.

[118] Thomas Monz, Daniel Nigg, Esteban A Martinez, Matthias F Brandl, Philipp Schindler, Richard Rines, Shannon X Wang, Isaac L Chuang, and Rainer Blatt. Realization of a scalable Shor algorithm. *Science*, 351(6277):1068–1070, 2016.

[119] J T Muhonen, A Laucht, S Simmons, J P Dehollain, R Kalra, F E Hudson, S Freer, K M Itoh, D N Jamieson, J C McCallum, A S Dzurak, and A Morello. Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. *Journal of Physics: Condensed Matter*, 27(15):154205, 2015.

[120] WJ Munro, DFV James, AG White, and Paul G Kwiat. Tomography and its role in quantum computation. *HP Laboratories Technical Report*, (53), 2001.

[121] Ravi Naik, Akel Hashim, David Santiago, and Irfan Siddiqi. AQT's superconducting quantum computing device from https://aqt.lbl.gov/.

[122] Charles Neill, Pedran Roushan, K Kechedzhi, Sergio Boixo, Sergei V Isakov, V Smelyanskiy, A Megrant, B Chiaro, A Dunsworth, K Arya, et al. A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science*, 360(6385):195–199, 2018.

[123] Erik Nielsen, Kenneth Rudinger, Timothy Proctor, Kevin Young, and Robin Blume-Kohout. Efficient flexible characterization of quantum processors with nested error models. *arXiv preprint arXiv:2103.02188*, 2021.

[124] Jeremy L O'Brien, Geoffrey J Pryde, Andrew G White, Timothy C Ralph, and David Branning. Demonstration of an all-optical quantum controlled-not gate. *Nature*, 426(6964):264–267, 2003.

[125] Ryo Okamoto, Holger F. Hofmann, Shigeki Takeuchi, and Keiji Sasaki. Demonstration of an optical quantum controlled-not gate without path interference. *Physical Review Letters*, 95:210506, 2005.

[126] TE Obrien, B Tarasinski, and L DiCarlo. Density-matrix simulation of small surface codes under current and projected experimental noise. *npj Quantum Information*, 3(1):1–8, 2017.

[127] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L Obrien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):1–7, 2014.

[128] Ch Piltz, Th Sriarunothai, AF Varón, and Ch Wunderlich. A trapped-ion-based quantum byte with 10- 5 next-neighbour cross-talk. *Nature communications*, 5(1):1–10, 2014.

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

[130] John Preskill. Quantum computing: pro and con. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):469–486, 1998.

[131] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout. What Randomized Benchmarking Actually Measures. *Physical Review Letters*, 119(13):130502, 2017.

[132] Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. Measuring the capabilities of quantum computers. *arXiv preprint arXiv:2008.11294*, 2020.

[133] Timothy J Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Direct randomized benchmarking for multi-qubit devices. *arXiv preprint arXiv:1807.07975*, 2018.

[134] Zbigniew Puchała, Łukasz Rudnicki, and Karol Życzkowski. Pauli semigroups and unistochastic quantum channels. *Physics Letters A*, 383(20):2376–2381, 2019.

[135] J. Qi and H. Khoon Ng. Randomized benchmarking does not measure average infidelity of gates. *ArXiv e-prints*, 2018.

[136] Isidor Isaac Rabi. Space quantization in a gyrating magnetic field. *Physical Review*, 51(8):652, 1937.

[137] Norman F Ramsey. A molecular beam resonance method with separated oscillating fields. *Physical Review*, 78(6):695, 1950.

[138] Andrey V. Rodionov, Andrzej Veitia, R. Barends, J. Kelly, Daniel Sank, J. Wenner, John M. Martinis, Robert L. Kosut, and Alexander N. Korotkov. Compressed sensing quantum process tomography for superconducting quantum gates. *Physical Review B*, 90:144504, 2014.

[139] MA Rol, Cornelis C Bultink, Thomas E OBrien, SR De Jong, Lukas S Theis, Xiang Fu, F Luthi, Raymond FL Vermeulen, JC De Sterke, Alessandro Bruno, et al. Restless tuneup of high-fidelity qubit gates. *Physical Review Applied*, 7(4):041001, 2017.

[140] Antoine Royer. Measurement of quantum states and the wigner function. *Foundations of physics*, 19(1):3–32, 1989.

[141] Kenneth Rudinger, Shelby Kimmel, Daniel Lobser, and Peter Maunz. Experimental demonstration of a cheap and accurate phase estimation. *Physical Review Letters*, 118(19):190502, 2017.

[142] Kenneth Rudinger, Timothy Proctor, Dylan Langharst, Mohan Sarovar, Kevin Young, and Robin Blume-Kohout. Probing context-dependent errors in quantum processors. *Physical Review X*, 9:021045, 2019.

[143] Łukasz Rudnicki, Zbigniew Puchała, and Karol Zyczkowski. Gauge invariant information concerning quantum channels. *Quantum*, 2:60, 2018.

[144] László Ruppert, Dániel Virosztek, and Katalin Hangos. Optimal parameter estimation of pauli channels. *Journal of Physics A: Mathematical and Theoretical*, 45(26):265305, 2012.

[145] Mary Beth Ruskai, Stanislaw Szarek, and Elisabeth Werner. An analysis of completely-positive trace-preserving maps on. *Linear Algebra and its Applications*, 347(1-3):159, 2002.

[146] Yuval R Sanders, Joel J Wallman, and Barry C Sanders. Bounding quantum gate error rate based on reported average fidelity. *New Journal of Physics*, 18(1):012002, 2015.

[147] Mohan Sarovar, Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. Detecting crosstalk errors in quantum information processors. *arXiv preprint arXiv:1908.09855*, 2019.

[148] Mohan Sarovar, Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. Detecting crosstalk errors in quantum information processors. *Quantum*, 4:321, 2020.

[149] Nicolas PD Sawaya, Mikhail Smelyanskiy, Jarrod R McClean, and Alán Aspuru-Guzik. Error sensitivity to environmental noise in quantum circuits for chemical state preparation. *Journal of chemical theory and computation*, 12(7):3097–3108, 2016.

[150] S. G. Schirmer, A. Kolli, and D. K. L. Oi. Experimental hamiltonian identification for controlled two-level systems. *Physical Review A*, 69:050306, 2004.

[151] Travis L Scholten, Yi-Kai Liu, Kevin Young, and Robin Blume-Kohout. Classifying single-qubit noise using machine learning. *arXiv preprint arXiv:1908.11762*, 2019.

[152] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Physical Review Letters*, 93(6):060302, 2016.

[153] Sarah Sheldon, Lev S Bishop, Easwar Magesan, Stefan Filipp, Jerry M Chow, and Jay M Gambetta. Characterizing errors on qubit operations via iterative randomized benchmarking. *Physical Review A*, 93(1):012301, 2016.

[154] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

[155] Matthias Steffen, Lieven M.K Vandersypen, and Isaac L Chuang. Simultaneous soft pulses applied at nearby frequencies. *Journal of Magnetic Resonance*, 146(2):369–374, 2000.

[156] M. Takita, A. D. Córcoles, E. Magesan, B. Abdo, M. Brink, A. Cross, J. M. Chow, and J. M. Gambetta. Demonstration of Weight-Four Parity Measurements in the Surface Code Architecture. *Physical Review Letters*, 117(21):210505, 2016.

[157] James Edward Tarlton. *Probing qubit memory errors at the* 10^-5 *level.* PhD thesis, Imperial College London, 2018.

[158] Honeywell team. Honeywell system model H0 from https://www.honeywell.com/us/en/company/ quantum/quantum-computer.

[159] IBM Q team. 5-qubit backend, “IBM Q 5 Yorktown backend specification V1.1.0,”. retrieved from https://quantum-computing.ibm.com, 2018.

[160] Quantum AI team and collaborators. Cirq, 2020.

[161] L. S. Theis, F. Motzoi, and F. K. Wilhelm. Simultaneous gates in frequency-crowded multilevel systems using fast, robust, analytic control shapes. *Physical Review A*, 93:012324, 2016.

[162] Géza Tóth, Witlef Wieczorek, David Gross, Roland Krischek, Christian Schwemmer, and Harald Weinfurter. Permutationally invariant quantum tomography. *Physical review letters*, 105(25):250403, 2010.

[163] T Van der Sar, ZH Wang, MS Blok, H Bernien, TH Taminiau, DM Toyli, DA Lidar, DD Awschalom, R Hanson, and VV Dobrovitski. Decoherence-protected quantum gates for a hybrid solid-state spin register. *Nature*, 484(7392):82–86, 2012.

[164] Lieven MK Vandersypen and Isaac L Chuang. Nmr techniques for quantum control and computation. *Reviews of modern physics*, 76(4):1037, 2005.

[165] K Vogel and H Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Physical Review A*, 40(5):2847, 1989.

[166] J. J. Wallman. Bounding experimental quantum error rates relative to fault-tolerant thresholds. *ArXiv e-prints*, 2015.

[167] J. J. Wallman. Randomized benchmarking with gate-dependent noise. *ArXiv e-prints*, 2017.

[168] Joel Wallman, Steven Flammia, and Ian Hincks. Quantum characterization, verification, and validation, 2018.

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

[170] Joel J Wallman. Bounding experimental quantum error rates relative to fault-tolerant thresholds. *arXiv preprint arXiv:1511.00727*, 2015.

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

[172] Joel J. Wallman, Marie Barnhill, and Joseph Emerson. Robust Characterization of Loss Rates. *Physical Review Letters*, 115(6):060501, 2015.

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

[174] Joel J. Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5):052325, 2016.

[175] Joel J. Wallman and Steven T. Flammia. Randomized benchmarking with confidence. *New Journal of Physics*, 16(10):103032, 2014.

[176] Stefan Weigert. How to determine a quantum state by measurements: the pauli problem for a particle with arbitrary potential. *Physical Review A*, 53(4):2078, 1996.

[177] Yaakov S Weinstein, Timothy F Havel, Joseph Emerson, Nicolas Boulant, Marcos Saraceno, Seth Lloyd, and David G Cory. Quantum process tomography of the quantum fourier transform. *The Journal of chemical physics*, 121(13):6117–6133, 2004.

[178] GAL White, CD Hill, FA Pollock, LCL Hollenberg, and Kavan Modi. Demonstration of non-markovian process characterisation and control on a quantum processor. *Nature Communications*, 11(1):1–10, 2020.

[179] Adam Winick, Joel J Wallman, and Joseph Emerson. Simulating and mitigating crosstalk. *arXiv preprint arXiv:2006.09596*, 2020.

[180] Christopher J Wood and Jay M Gambetta. Quantification and characterization of leakage errors. *Physical Review A*, 97(3):032306, 2018.

[181] T Xia, M Lichtman, K Maller, AW Carr, MJ Piotrowicz, L Isenhower, and M Saffman. Randomized benchmarking of single-qubit gates in a 2d array of neutral-atom qubits. *Physical Review Letters*, 114(10):100503, 2015.