Randomized Benchmarking for Noise Characterization

Sumit Sijher

(Dated: 8 August, 2018)

This paper serves to provide an overview of the standard Randomized Benchmarking protocol which is used to obtain experimental estimates of the average error associated to a set of gates that form a 2-design such as the Clifford group. As motivation for the protocol, we discuss traditional noise characterization methods relying on Quantum State and Process Tomography highlighting their practical inefficiency and sensitivities to state preparation and measurement errors. We keep the presentation mostly self contained and include a discussion of the relevant definitions and constructions needed to understand the Randomized Benchmarking protocol. In addition, we remark on various extensions and specializations that have been inspired by the standard protocol in order to characterize other essential features of noise that can effect general quantum information processing architectures.

I. INTRODUCTION

Although ideal faultless quantum information processing is an interesting idea to entertain, it seems inevitable that errors and unwanted noise will be present in any physical platform. In pursuit of this goal, it is necessary to have methods for understanding the nature of noise present in the system. In principle, traditional methods like quantum state and process tomography (QST/QPT) [1] offer a complete characterization of states and processes in a system. However, such methods typically require resources that scale exponentially in the system size, and are prone to state preparation and measurement (SPAM) errors. Therefore quantum tomography is neither efficient nor robust in practice.

In contrast to the costly and imperfect approach of quantum tomography, Randomized Benchmarking (RB) protocols instead offer partial characterization of noise processes that are both efficient and robust to SPAM errors. The original context of noise for RB protocols [2] is to diagnose the nature of imperfect gates in a quantum computer. In this regard, RB protocols aim to determine the average gate error associated to a group of unitary gates referred to as the Clifford group, which play a central role in quantum computation. These protocols suggest applying various random sequences of faulty gates to some input state in such a way that certain average output statistics yield information about the noise associated to the gates. In particular, the data gathered from these experiments are fitted to an exponential decay model in order to infer the decay parameter which is related to the fidelity of the underlying noise.

In addition to determining average gate fidelities for the Clifford group, RB protocols have been developed which serve to characterize the noise associated to a single gate of interest in the Clifford group [3], and also gates that lie outside Clifford group as well [4, 5]. Moreover, RB inspired protocols have been developed to characterize particular properties of the noise such system loss [6], coherence or unitarity [7], and addressability [8].

The structure of this paper is outlined as follows. In Section II, we briefly review the methods behind Quantum State and Process Tomography with a discussion of its main pitfalls regarding resource complexity and sensitivity to SPAM errors. In Section III, we define the relevant notions of fidelity and twirling that are necessary to understand what the relevant figure of merit RB seeks to determine and how it aims to do so. Having developed this definitions and constructions, Section IV will describe the main RB protocol and provide an analysis and derivation. In Section V, we conclude with a discussion of various ways that RB methods have been utilized to go beyond the standard framework to characterize other features of the noise. For the unacquainted reader, included at the end of this work is an Appendix which gives some general definitions and constructions in quantum information theory that are used throughout this work.

II. QUANTUM STATE AND PROCESS TOMOGRAPHY

To help motivate some of the merits of Randomized Benchmarking, we will start with a brief description of quantum state and process tomography and illuminate some of the pitfalls and experimental challenges for implementing these methods in practice. We refer the reader to the Appendix for details regarding standard definitions and constructions in the theory of quantum information that will be used throughout.

In principle, quantum state and process tomography offers a standard method to experimentally characterize general states and process involved in quantum information processing. In practice however, the complete determination of either states or process involves a number of experiments that grow exponentially in terms of the size of the system.

For illustrative purposes, consider a Hilbert space $\mathcal{H}$ consisting of n qubits with dimension $d = 2^n$. A general *quantum state* of $\mathcal{H}$ is an operator $\rho \in D(\mathcal{H})$ which by definition is Hermitian, positive semi-definite, and has unit trace. A *quantum operation* is a linear map $\mathcal{E} : D(\mathcal{H}) \to D(\mathcal{H})$ between states that is *completely positive* (CP) and *trace non-increasing*. A map $\mathcal{E}$ is *trace preserving* if $Tr(\mathcal{E}(\rho)) = Tr(\rho)$ for all $\rho \in D(\mathcal{H})$. A completely positive trace preserving (CPTP) map is often referred to as a *quantum channel* in the literature.

States and operations can be described more explicitly by choosing a suitable operator basis $\mathcal{B}$ of the space of bounded operators $B(\mathcal{H})$. Throughout, we will take this basis to be the n-qubit Pauli group $\mathcal{B} := \{P_i\}_{i=1}^{d^2}$, which is indexed by d^2 many elements.

A state can be expanded with respect to this basis as

$$\rho = \sum_{i=1}^{d^2} Tr(P_i\rho)P_i, \tag{1}$$

where the expansion coefficients $Tr(P_i\rho)$ carry the operational interpretation of being expectation values of the observables P_i by virtue of the Pauli group consisting of Hermitian operators. This is precisely the experimental prescription for standard QST: for each $P_i \in \mathcal{B}$, estimate the expectation value $Tr(P_i\rho)$ on some state of interest ρ. Although seemingly straightforward, note that there are a total of d^2 such observables, which for a n-qubit system means that a total of $d^2 = 2^{2n}$ experiments would be necessary to completely characterize the state ρ as expressed in Equation 1. As this quantity is exponential in n, the resources for such a task quickly become impractical.

Likewise, a quantum operation also affords an expansion in terms of the basis elements given as

$$\mathcal{E}(\rho) = \sum_{m,n} \chi_{mn} P_m \rho P_n, \tag{2}$$

The χ-matrix comprised of these expansion coefficients, $[\chi]_{m,n} := \chi_{mn}$, completely characterizes the action of a quantum operation $\mathcal{E}$. Note that the χ-matrix is a $d^2 \times d^2$ matrix, but due to the conditions of complete positivity and trace preservation, may have $d^4 - d^2$ independent parameters. Regardless, this quantity still scales exponentially in the number of qubits n.

Quantum Process Tomography leverages the method of QST to experimentally determine the action of some "black-box" quantum operation $\mathcal{E}$. This is done by choosing a particular set of input states $\{\rho_j\}_{j=1}^{d^2}$ which span $D(\mathcal{H})$, applying the operation $\mathcal{E}$ to each state, and then performing QST as described above to each output state $\mathcal{E}(\rho_j)$. Again, a total of d^2 observables much be determined for each output state, which means a total of $d^4 = 2^{4n}$ different experiments would be required to characterize the unknown operation $\mathcal{E}$. Moreover, we remark that the experimental data gathered from QPT must be classically post-processed via linear inversion of a $d^4 \times d^4$ matrix in order to calculate estimates all the relevant parameters of the χ-matrix. Hence, QPT quickly becomes infeasible in practice.

Despite the exponential experimental resource requirements, both QST and QPT methods are further complicated by errors do to imperfect state preparation and measurement (SPAM). Measurement errors may arise in the estimation of any of the expectation values $Tr(P_i\rho)$ for the observables $P_i \in \mathcal{B}$. Additionally, for QPT, imperfect state preparation of the input states $\{\rho_j\}$ may compound the error even further. If the main objective of these protocols is to determine some unknown operation $\mathcal{E}$, but noise is present in either state preparation or measurement, then a priori it may not be possible to understand the dynamics of $\mathcal{E}$ without conflating noise arising from SPAM. This is what it is meant in saying that QST and QPT are not robust to SPAM, which consequently forms an inherent challenge and shortcoming of these methods.

In the following sections we will describe the standard Randomized Benchmarking protocol, which overcomes the limitations of QST/QPT by being both efficient in practice and robust to SPAM errors. This however comes at a trade off, and seems to require forfeiting a complete characterization of a quantum operation $\mathcal{E}$, in exchange for only partial information that yields certain characterizing figures of merit.

III. FIDELITY AND TWIRLING

Before presenting the RB protocol, we will discuss in this section the notion of *fidelity* between quantum states and operations. These definitions will form the basis of the main figure of merit provided by RB protocols, which is the average fidelity of a group of operations known as Clifford gates. Moreover, we will discuss certain essential properties of the fidelity and how it relates to a superoperator construct known as the *twirl* of a quantum operation, which will form the basis of the RB protocol.

A. Fidelity of States and Operations

The *fidelity* between two states $\rho, \sigma \in D(\mathcal{H})$ is defined as the function

$$F(\rho, \sigma) := Tr\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2, \tag{3}$$

and when one of the states is pure, say $\rho = |\psi\rangle\langle\psi|$, this takes the simpler form

$$F(|\psi\rangle\langle\psi|, \sigma) = Tr(|\psi\rangle\langle\psi|\sigma) = \langle\psi|\sigma|\psi\rangle, \tag{4}$$

which has the geometric interpretation of describing the "overlap" between the state $|\psi\rangle\langle\psi|$ and σ. The fidelity function is bounded as $0 \leq F(\rho, \sigma) \leq 1$, and has the property that $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$.

Using this definition of state fidelity, we can define the fidelity between two quantum operations $\mathcal{E}_1$ and $\mathcal{E}_2$ on a particular state ρ as

$$F_{\mathcal{E}_1,\mathcal{E}_2}(\rho) := F(\mathcal{E}_1(\rho), \mathcal{E}_2(\rho)). \tag{5}$$

Now suppose that the input state is pure $\rho = |\psi\rangle\langle\psi|$ and one of the operations under consideration is the identity operation $\mathcal{I} : D(\mathcal{H}) \to D(\mathcal{H})$. Then we can write

$$F_{\mathcal{I},\mathcal{E}}(|\psi\rangle\langle\psi|) = Tr(|\psi\rangle\langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)), \tag{6}$$

and interpret this quantity as a survival probability which effectively measure how much the operation $\mathcal{E}$ "distorts" the input state $|\psi\rangle\langle\psi|$. In what follows, we will ultimately be concerned with characterizing some noise operation $\mathcal{E}$ which in the ideal case should be noise free: $\mathcal{E} = \mathcal{I}$. Hence, we will mostly be interested in comparing some operation $\mathcal{E}$ to the identity $\mathcal{I}$, and focus our attention on the expression given in Equation 6.

Note that the fidelity functions in Equations 5 and 6 depend on the input state in the argument, but to truly compare two different operations we should account for

their actions on all possible input states. In this regard, we define the *average fidelity* of an operation $\mathcal{E}$ as

$$F_{avg}(\mathcal{E}) := \int d\psi F_{\mathcal{I},\mathcal{E}}(|\psi\rangle\langle\psi|), \tag{7}$$

where the integral is taken over all pure states via the *Fubini-Study measure* $d\psi$. In this way, $F_{avg}(\mathcal{E})$ effectively compares the average distortion of all input states $|\psi\rangle\langle\psi|$ to their output states. Alternatively, since the unitary group $\mathcal{U}(\mathcal{H})$ acts transitively on the space $\mathcal{H}$, we can alternatively define the average fidelity as

$$F_{avg}(\mathcal{E}) := \int dU F_{\mathcal{I},\mathcal{E}}(U|\psi\rangle\langle\psi|U^\dagger), \tag{8}$$

where now the integral is taken over all unitary operators $U \in \mathcal{U}(\mathcal{H})$ via the *Haar* measure dU, and $|\psi\rangle\langle\psi|$ is some fixed arbitrary pure state.

One essential property of the average fidelity defined above is that it is invariant under unitary conjugation. To make this more precise, for a unitary operation $U \in \mathcal{U}(\mathcal{H})$ denote by $\mathcal{U}$ the channel given by $\mathcal{U}(\rho) = U\rho U^\dagger$, and consider the fidelity of the conjugation composition $\mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U}$ which can be expressed as

$$\begin{aligned} F_{\mathcal{I},\mathcal{U}^\dagger\circ\mathcal{E}\circ\mathcal{U}}(|\psi\rangle\langle\psi|) &= Tr(|\psi\rangle\langle\psi|\mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U}(|\psi\rangle\langle\psi|)) \\ &= Tr(|\psi\rangle\langle\psi|U^\dagger\mathcal{E}(U|\psi\rangle\langle\psi|U^\dagger)U) \\ &= Tr(U|\psi\rangle\langle\psi|U^\dagger\mathcal{E}(U|\psi\rangle\langle\psi|U^\dagger)) \\ &= F_{\mathcal{I},\mathcal{E}}(|\phi\rangle\langle\phi|), \end{aligned}$$

where we have defined $|\phi\rangle := U|\psi\rangle$. Then by averaging both sides over all pure states it is observed that

$$F_{avg}(\mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U}) = F_{avg}(\mathcal{E}), \tag{9}$$

implying that the average fidelity of the operation $\mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U}$ is the same as the average fidelity of $\mathcal{E}$. This particular equivalence is an important insight and will be exploited in the RB protocol.

B. Channel Twirling

Consider now some subgroup $\mathcal{G} \subseteq \mathcal{U}(\mathcal{H})$ of the unitary group. Motivated by the unitary conjugation construction of a map $\mathcal{E}$ described above, we define the $\mathcal{G}$-*twirl* of a map $\mathcal{E}$ as

$$\mathcal{E}^{\mathcal{G}} := \frac{1}{|\mathcal{G}|}\sum_{\mathcal{U}\in\mathcal{G}} \mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U}, \tag{10}$$

which is just the uniform average of the maps $\mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U}$ over all $\mathcal{U} \in \mathcal{G}$. Note that we implicitly assumed that the subgroup $\mathcal{G}$ is discrete or finite and hence the summation over group elements. In the case that the subgroup $\mathcal{G}$ is not discrete, a suitable integral must be taken in defining the averaged operation. For instance, suppose that the chosen group is itself the full unitary group: $\mathcal{G} = \mathcal{U}(\mathcal{H})$. Then the full $\mathcal{U}(\mathcal{H})$-twirl is given by

$$\mathcal{E}^{\mathcal{U}(\mathcal{H})} := \int_{\mathcal{U}\in\mathcal{U}(\mathcal{H})} \mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U} d\mathcal{U}, \tag{11}$$

where the integral is performed uniformly over all $\mathcal{U} \in \mathcal{U}(\mathcal{H})$ via the Harr measure $d\mathcal{U}$. In this case, it can be shown that the action of the fully twirled channel $\mathcal{E}^{\mathcal{U}(\mathcal{H})}$ takes a special form:

$$\mathcal{E}^{\mathcal{U}(\mathcal{H})} := p\rho + (1-p)\mathbb{I} = \mathcal{D}_p(\rho), \tag{12}$$

which can be recognized as the *depolarizing channel* with depolarizing parameter p. This channel has the operational interpretation that with probability p does nothing to the input state (performs the identity channel), and with probability $1-p$ gives the maximally mixed state $\mathbb{I} \in D(\mathcal{H})$.

As shown above, the fact that the average fidelity of a map $\mathcal{E}$ is invariant under unitary conjugation also implies that the average fidelity of the twirled channel $\mathcal{E}^{\mathcal{G}}$ is also invariant:

$$F_{avg}(\mathcal{E}^{\mathcal{G}}) = F_{avg}(\mathcal{E}). \tag{13}$$

In the case of the full $\mathcal{U}(\mathcal{H})$-twirl, we can calculate the average fidelity using Equation 12 as

$$F_{avg}(\mathcal{E}^{\mathcal{U}(\mathcal{H})}) = F_{avg}(\mathcal{D}_p) = p + \frac{1-p}{d}, \tag{14}$$

which is given in terms of the depolarizing parameter p and dimension d of the Hilbert space $\mathcal{H}$. Thus, by performing a full twirl of an arbitrary channel $\mathcal{E}$, we can determine $F_{avg}(\mathcal{E})$ by determining $F_{avg}(\mathcal{E}^{\mathcal{U}(\mathcal{H})})$.

A natural question then is how can one operationally implement the full twirl $\mathcal{E}^{\mathcal{U}(\mathcal{H})}$? After all, such a channel involves an average over infinitely many unitary operations. Fortunately, this is not necessary thanks to the existence of special algebraic structures referred to as unitary 2-designs, which allow the realization of this uniform average over infinitely many elements through finite means. Although several equivalent definitions exist [9], it will suffice here to simply define a *unitary 2-design* as a subgroup $\mathcal{G} \subset \mathcal{U}(\mathcal{H})$ with the property that $\mathcal{E}^{\mathcal{G}} = \mathcal{E}^{\mathcal{U}(\mathcal{H})}$ for all channels $\mathcal{E}$. Therefore, if $\mathcal{G}$ is a 2-design,

$$\mathcal{E}^{\mathcal{G}} = \frac{1}{|\mathcal{G}|}\sum_{\mathcal{U}\in\mathcal{G}} \mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U} = \int_{\mathcal{U}(\mathcal{H})} \mathcal{U}^\dagger \circ \mathcal{E} \circ \mathcal{U} d\mathcal{U} = \mathcal{E}^{\mathcal{U}(\mathcal{H})} \tag{15}$$

Hence, twirling over a 2-design group $\mathcal{G}$ is operationally equivalent to twirling over the full unitary group implying that $\mathcal{E}^{\mathcal{G}} = \mathcal{D}_p$ for some p. This is beneficial when $\mathcal{G}$ is itself a smaller finite group providing a more practical means to implement the desired twirled channel.

The canonical example of a 2-design is the *Clifford group* $\mathcal{C}_n$ on n qubits, which is defined as the normalizer of the Pauli group $\mathcal{P}_n$:

$$\mathcal{C}_n := \{U \in \mathcal{U}(\mathcal{H}) : UPU^\dagger \in \mathcal{P}_n, \forall P \in \mathcal{P}_n\} \tag{16}$$

The Clifford group is ubiquitous throughout the theory of quantum computation, serving a central role in many tasks including quantum error correction and fault tolerance. We merely remark here, that by the *Gottesman-Knill Theorem* [10], the action of Clifford gates can be efficiently simulated classically unlike the general simulation of arbitrary unitary gates. This property is essential in the feasibility of the standard Randomized Benchmarking protocol, where it is necessary to track the ideal

dynamics of sequences of Clifford gates in order to calculate the appropriate inverse gate for the desired motion reversal as discussed below.

IV. RANDOMIZED BENCHMARKING

Having developed the necessary background material in the previous sections, we can now describe the standard RB protocol. Before doing so, we make clear what it is that RB intends to estimate. In practice a unitary operation $\mathcal{U}$ may not be implemented ideally and instead a noisy operation $\tilde{\mathcal{U}}$ is implemented instead. Without loss of generality, we represent the noisy implementation as $\tilde{\mathcal{U}} = \mathcal{E}_{\mathcal{U}} \circ \mathcal{U}$ and allow $\mathcal{E}_{\mathcal{U}}$ to be an arbitrary CPTP map which describes the noise associated to $\mathcal{U}$. In the ideal, noise-free case $\mathcal{E}_{\mathcal{U}} = \mathcal{I}$ is just the identity. Consider the Clifford group $\mathcal{C}$, and for each $\mathcal{U} \in \mathcal{C}$ let $\tilde{\mathcal{U}} = \mathcal{E}_{\mathcal{U}} \circ \mathcal{U}$ be the noisy implementation of the Clifford gate. Define the *average error map*

$$\overline{\mathcal{E}} := \frac{1}{|\mathcal{C}_n|} \sum_{\mathcal{U} \in \mathcal{C}_n} \mathcal{E}_{\mathcal{U}}. \tag{17}$$

Then the RB protocol offers a means to estimate the average fidelity $F_{avg}(\overline{\mathcal{E}})$ of the average error map associated to the Clifford group by first obtaining an estimate of the depolarizing parameter p. This is done via the relationship:

$$F_{avg}(\overline{E}) = p + \frac{(1-p)}{d} \tag{18}$$

In particular, we will only concern ourselves here with the *zeroth-order model*, which makes the rather unrealistic assumption that the errors $\mathcal{E}_{\mathcal{U}}$ are both gate and time independent. This implies that for all $\mathcal{U} \in \mathcal{C}_n$, there exists a unique CPTP map $\mathcal{E}$ such that $\mathcal{E}_{\mathcal{U}} = \mathcal{E}$. In this case, the average error map is given by $\overline{\mathcal{E}} = \mathcal{E}$. Perturbative approaches can be applied to relax these assumptions of gate and time independence [2, 11].

First, let us fix some notation for convenience. Let $\mathbb{N}_N := \{1, \ldots, N\}$ denote the set of the first N natural numbers, and $\mathbb{N}_N^m$ the m-fold Cartesian product of $\mathbb{N}_N$ with itself. An ordered sequence $\vec{k}_m := (k_1, \ldots, k_m) \in \mathbb{N}_{|\mathcal{C}|}^m$ will index a collection of noisy gates $\tilde{\mathcal{U}}_{k_i}$ corresponding to ideal Clifford gates $\mathcal{U}_{k_i} \in \mathcal{C}$ and serve to define the ordered gate composition

$$\tilde{\mathcal{U}}_{\vec{k}_m} := \bigcirc_{i=1}^m \tilde{\mathcal{U}}_{k_i} = \tilde{\mathcal{U}}_{k_m} \circ \cdots \circ \tilde{\mathcal{U}}_{k_1} \tag{19}$$

and its corresponding ideal gate sequence $\mathcal{U}_{\vec{k}_m}$ In addition, we define the ideal inverse gate associated to the sequence as

$$\mathcal{U}_{\vec{k}_m}^\dagger := \bigcirc_{i=m}^1 \mathcal{U}_{k_i} = \mathcal{U}_{k_1}^\dagger \circ \cdots \circ \mathcal{U}_{k_m}^\dagger, \tag{20}$$

and its noisy implementation as $\tilde{\mathcal{U}}_{\vec{k}_m}^\dagger = \mathcal{E} \circ \mathcal{U}_{\vec{k}_m}^\dagger$. Then the full sequence corresponding to the noisy motion reversal is given as

$$\tilde{S}_{\vec{k}_m} := \tilde{\mathcal{U}}_{\vec{k}_m}^\dagger \circ \tilde{\mathcal{U}}_{\vec{k}_m}, \tag{21}$$

where in the ideal noise-free case $S_{\vec{k}_m} = \mathcal{I}$.

For some positive integer m, let $K_m := \{\vec{k}_m^{(s)}\}_{s=1}^{|K_m|}$ be a set of $|K_m|$ sequences, each of length m, where each component of a given sequence $\vec{k}_m^{(s)} \in \mathbb{N}_{|\mathcal{C}|}^m$ is chosen independently and uniformly at random. This ensures that the corresponding gates will be sampled from $\mathcal{C}$ independently and uniformly at random, which is important for faithfully constructing the twirled channel. Now define the average sequence

$$\tilde{S}_{K_m}^{avg} := \frac{1}{|K_m|} \sum_{\vec{k}_m \in K_m} \tilde{S}_{\vec{k}_m} \tag{22}$$

The size of the set $|K_m|$ determines the accuracy of the fidelity estimate in that averaging over sequences in $|K_m|$ yields an approximate twirl over the full Clifford group, which is what the RB protocol merely aims to achieve. It is in the limiting case when the average is taken over all possible Clifford gate sequences that the full 2-design twirl is realized [12].

In short, the RB protocol seeks to obtain an estimate of the depolarizing parameter p, which in turn can be related to the average fidelity of the noise channel as in Equation . To accomplish this RB aims to estimate the survival probability of some state $|\psi\rangle\langle\psi|$ under the action of the averaged motion reversal sequence described in Equation 22. This requires the preparation of some arbitrary experimental initial state $\rho = |\psi\rangle\langle\psi| \in D(\mathcal{H})$, and a final experimental measurement which we will take to be given by the POVM $E := \{|\psi\rangle\langle\psi|, \mathbb{I} - |\psi\rangle\langle\psi|\}$. However, these primitives may themselves be noisy, which we represent as some noisy state preparation $\tilde{\rho} \in D(\mathcal{H})$ and noisy measurement given by a POVM $\tilde{E}$. In this regard, denote the survival probability of a successful motion reversal using the noisy operations for a given sequence $\vec{k}_m^{(s)} \in K_m$ as

$$Q_{\vec{k}_m^{(s)}} = Tr(\tilde{E} \tilde{S}_{\vec{k}_m^{(s)}}(\tilde{\rho})). \tag{23}$$

Moreover, define the survival probability under the averaged sequence of Equation 22 as

$$Q_{K_m} = Tr(\tilde{E} \tilde{S}_{K_m}^{avg}(\tilde{\rho})). \tag{24}$$

The protocol proceeds as follows.

A. Randomized Benchmarking Protocol

1. Fix a positive integer m
2. Sample a set $K_m := \{\vec{k}_m^{(s)}\}_{s=1}^{|K_m|}$ of sequences $\vec{k}_m^{(s)} \in \mathbb{N}_{|\mathcal{C}|}^m$, where each component of $\vec{k}_m^{(s)}$ is chosen independently and uniformly at random.
3. For each sequence $\vec{k}_m^{(s)} \in K_m$, apply the gate sequence $\tilde{\mathcal{U}}_{\vec{k}_m^{(s)}}$ to the initial state $\tilde{\rho}$, followed by the inversion gate, and a measurement of $\tilde{E}$ in order to estimate

$$Q_{\vec{k}_m^{(s)}} = Tr(\tilde{E} \tilde{S}_{\vec{k}_m^{(s)}}(\tilde{\rho})) \tag{25}$$

to a desired precision by repeating the same gate sequence given by $\tilde{S}_{\vec{k}_m^{(s)}}$.

4. Estimate the true survival probability by averaging over all sequences in K_m with

$$Q_{K_m} = \frac{1}{|K_m|} \sum_{s=1}^{|K_m|} Q_{\vec{k}_m^{(s)}} \tag{26}$$

5. Repeat steps $1-4$ for increasing values of m and fit the data to the decay curve as a function of sequence length

$$f(m) := Ap^m + B \tag{27}$$

in order to estimate the decay parameter p, where the constant A and B carry information about the SPAM as defined below in Equation 33.

In order to prove the correctness of the RB protocol, we will calculate what the true survival probability denoted by Q_m ought to be when the randomization over gates realizes the full Clifford twirl. The main analytical claim is that, as a function of m, the survival probability is given by a decay curve of the form expressed in Equation 27. On the other hand, the main statistical claim is that sampling over just a subset of all possible sequences given by $K_m \subset \mathbb{N}_{|\mathcal{C}|}^m$ yields an approximation of Q_m, such that in the limit as $K_m \to \mathbb{N}_{|\mathcal{C}|}^m$, the survival probability converges as $Q_{K_m} \to Q_m$. We focus our attention on the former claim.

Consider the noisy motion reversal sequence as given in Equation 21 and factor each noisy gate in terms of the ideal gate followed by the noise channel $\mathcal{E}$:

$$\tilde{S}_{\vec{k}_m} = \mathcal{E} \circ \mathcal{U}_{\vec{k}_m}^\dagger \circ \mathcal{E} \circ \mathcal{U}_{k_m} \circ \cdots \circ \mathcal{E} \circ \mathcal{U}_{k_1}. \tag{28}$$

Now define for each index $1 \leq i \leq m$ the ideal gates $\mathcal{V}_{k_i} = \mathcal{U}_{k_i} \circ \cdots \circ \mathcal{U}_{k_1}$, which allows us to re express the sequence as

$$\tilde{S}_{\vec{k}_m} = \mathcal{E} \bigcirc_{i=1}^m \left(\mathcal{V}_{k_i}^\dagger \circ \mathcal{E} \circ \mathcal{V}_{k_i} \right). \tag{29}$$

Then averaging over all possible Clifford gates for sequences $\vec{k}_m \in \mathbb{N}_{|\mathcal{C}|}^m$ results in the average sequence

$$\begin{aligned} \tilde{S}_{\mathbb{N}_{|\mathcal{C}|}^m}^{avg} &= \frac{1}{|\mathcal{C}|^m} \sum_{\vec{k}_m \in \mathbb{N}_{|\mathcal{C}|}^m} \tilde{S}_{\vec{k}_m} \\ &= \mathcal{E} \bigcirc_{i=1}^m \left(\frac{1}{|\mathcal{C}|} \sum_{k_i \in \mathbb{N}_{|\mathcal{C}|}} \mathcal{V}_{k_i}^\dagger \circ \mathcal{E} \circ \mathcal{V}_{k_i} \right) \\ &= \mathcal{E} \bigcirc_{i=1}^m \mathcal{E}^{\mathcal{C}}, \end{aligned}$$

where in the second identity we equivalently average each index $k_i \in \mathbb{N}_{|\mathcal{C}|}$ independently which realizes the full Clifford twirl of the error $\mathcal{E}$ in the last line. Then as a consequence of the 2-design condition, this results in a composition of depolarizing channels. Thus, the full averaged sequence becomes

$$\tilde{S}_{\mathbb{N}_{|\mathcal{C}|}^m}^{avg} = \mathcal{E} \bigcirc_{i=1}^m \mathcal{D}_p. \tag{30}$$

Using Equation 12, the action of the m-fold composition of the depolarizing channel is given as

$$\bigcirc_{i=1}^m \mathcal{D}_p(\rho) = p^m \rho + (1-p^m)\mathbb{I}. \tag{31}$$

Therefore true survival probability for sequences of length m can be calculated as

$$\begin{aligned} Q_m &= Tr(\tilde{E} \tilde{S}_{\mathbb{N}_{|\mathcal{C}|}^m}^{avg}(\tilde{\rho})) \\ &= Tr(\tilde{E}\mathcal{E} \bigcirc_{i=1}^m \mathcal{D}_p(\tilde{\rho})) \\ &= Tr(\tilde{E}\mathcal{E}(p^m \tilde{\rho} + (1-p^m)\mathbb{I})). \end{aligned}$$

This can be expressed as a function of m as the exponential decay curve

$$f(m) = Tr(\tilde{E}\mathcal{E}(\tilde{\rho} - \mathbb{I}))p^m + Tr(\tilde{E}\mathcal{E}(\mathbb{I})) = Ap^m + B, \tag{32}$$

where we have introduced the SPAM constants

$$A := Tr(\tilde{E}\mathcal{E}(\tilde{\rho} - \mathbb{I})) \tag{33}$$

$$B := Tr(\tilde{E}\mathcal{E}(\mathbb{I})). \tag{34}$$

V. CONCLUSION

In summary, this work has described the standard Randomized Benchmarking protocol which can be used to characterize the average noise associated to a set of gates. The main merits of this protocol is its efficiency and robustness to errors due to imperfect state preparation and measurement, which are issues that plague other traditional methods for characterizing noise such as Quantum State and Process Tomography. In particular, the RB protocol provides an estimate of the average fidelity associated to a group of unitary gates that form a 2-design; the canonical example being the Clifford group. This is accomplished by estimating the survival probability of increasingly long sequences of gates that effectively perform a motion reversal experiment. By averaging the statistics accordingly, the data can be fitted to an exponential decay curve in order to extract the decay parameter which in turn is related to the average fidelity of the noise.

The analysis given in this work corresponds to the zeroth-order model of [2] which assumed both gate and time independent errors for an otherwise arbitrary Markovian error model. Higher order models can be analyzed to relax these assumptions as in [2, 12]. These results involve utilizing perturbative expansions, which yield a more intricate analysis albeit with a similar essence. Furthermore, work has been done to account for error models that may violate some of these primary assumptions such as non-Markovian noise [13].

In addition, work has been done which extends the benchmarking capabilities outside the 2-design regime. Such protocols are similar in spirit to traditional RB protocols and are able to benchmark individual gates of interest as in *Interleaved* RB [3], and also gates not belonging to the Clifford group allowing an analysis for characterizing universal gate sets [4, 5]. Further more, specific RB inspired protocols have been designed to benchmark other important characteristics of noise such as system loss [6], coherence or unitarity [7], and addressability [8].

Lastly, we mention how RB techniques can be used to help mitigate certain types of noise that may be more catastrophic in practice, which may allow for an increase of signal-to-noise. One prime example is the technique referred to as *Randomized Compiling* [14]. This work aims to use similar randomized approaches to compile various circuits into a particular form in order to effectively tailor the general noise present in the system to a more manageable form of stochastic Pauli noise.

Noise present in real-world implementations of quantum information processing tasks pose a challenging obstacle in accomplishing the desired tasks. Although this is still an active area of research, Randomized Benchmarking methods will surely become an essential paradigm of noise characterization and diagnostics for noise present in general quantum information process architectures.

Appendix: Quantum States, Operations, and the χ-matrix Representation

This appendix will serve to give a brief mathematical review of the standard formalism for quantum operations and describe some useful representations. This formalism will serve to describe the general transformations of quantum systems, which includes both the intended dynamics and the unintended dynamics in the form of noise. Let $\mathcal{H}$ be a Hilbert space representing the state space of a quantum system of interest, and $B(\mathcal{H})$ the set of bounded linear operators on $\mathcal{H}$. Throughout this work we will assume that $\mathcal{H}$ has finite dimension d so that $\mathcal{H} \cong \mathbb{C}^d$. A general quantum state ρ, referred to as a *density operator*, is an element of $D(\mathcal{H}) \subset B(\mathcal{H})$ corresponding to positive semi-definite, Hermitian operators having unit trace. We say that ρ is *normalized* if it has unit trace $Tr(\rho) = 1$, and is *unnormalized* if $Tr(\rho) < 1$. A state ρ is *pure* if it has rank 1 so that $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{H}$ having unit norm. More generally, a state is *mixed* if it is a convex combination of pure states: $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, for positive real numbers p_i such that $\sum_i p_i = 1$. The convex structure of $D(\mathcal{H})$ allows us to also consider an *ensemble* of states given as $\rho = \sum_i p_i \rho_i$ for arbitrary states $\rho_i \in D(\mathcal{H})$.

A *quantum operation* or *channel* $\mathcal{E}$ which maps input states of $\mathcal{H}$ to output states of $\mathcal{H}'$ is given by a linear map

$$\mathcal{E} : D(\mathcal{H}) \to D(\mathcal{H}')$$
$$\rho \mapsto \rho' := \frac{\mathcal{E}(\rho)}{Tr(\mathcal{E}(\rho))}$$

Here, the output state ρ' has been normalized by the factor $Tr(\mathcal{E}(\rho))$ to ensure that $\rho' \in D(\mathcal{H}')$ has unit trace. In general, the quantum operation $\mathcal{E}$ must be both completely positive and trace decreasing. *Trace decreasing* means that $Tr(\mathcal{E}(\rho)) \leq Tr(\rho) = 1$ for all $\rho \in D(\mathcal{H})$. As a special case, $\mathcal{E}$ is said to be *trace preserving* (TP) if $Tr(\mathcal{E}(\rho)) = 1$ for all $\rho \in D(\mathcal{H})$. The map $\mathcal{E}$ is *completely positive* (CP) if for any state $\sigma \in D(\mathcal{H}) \otimes D(\mathcal{H}_R)$ belonging to a larger space formed by considering an additional reference system $\mathcal{H}_R$, the output state given by $(\mathcal{E} \otimes \mathcal{I}_R)(\sigma)$ that results from applying $\mathcal{E}$ to $D(\mathcal{H})$ and the identity $\mathcal{I}_R$ on the rest of the space $D(\mathcal{H}_R)$ results in a positive operator. A map that is both completely positive and trace preserving will be abbreviated as CPTP. Throughout, we will restrict out attention to CP maps $\mathcal{E} : D(\mathcal{H}) \to D(\mathcal{H})$ from the space $D(\mathcal{H})$ to itself. Moreover, we will be mostly concerned with trace preserving maps.

The action of a CP map $\mathcal{E}$ on a state $\rho \in D(\mathcal{H})$ can be represented as

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{A.1}$$

for a set of linear operators $\{K_i\}_i \subset B(\mathcal{H})$ acting on $\mathcal{H}$. The K_i are referred to as *Kraus operators* or *decomposition operators*, and this representation of $\mathcal{E}$ is referred to as a *Kraus decomposition* or *operator-sum decomposition*. Although such a decomposition completely determines the action of $\mathcal{E}$, the choice of Kraus operators may not be unique.

The trace decreasing criterion for a quantum operation $\mathcal{E}$ implies that its Kraus operators must satisfy

$$\sum_i K_i^\dagger K_i \leq I, \tag{A.2}$$

For the case of a trace preserving map equality holds giving the completeness condition:

$$\sum_i K_i^\dagger K_i = I, \tag{A.3}$$

A CP map $\mathcal{E}$ is said to be *pure* if its Kraus decomposition consists of a single Kraus operator K so that $\mathcal{E}(\rho) = K\rho K^\dagger$. It then follows that a pure CPTP map is given by a unitary operator, $\mathcal{E}(\rho) = U\rho U^\dagger$, since the completeness condition implies $U^\dagger U = I$.

Another useful representation of CP maps allows us to interpret a map $\mathcal{E}$ as a matrix, referred to as the χ-matrix representation. This is possible given the linearity of the operation, but to give an explicit matrix representation requires choosing a basis for the space $B(\mathcal{H})$. Here, we will choose the basis to be the qudit Pauli group $\mathcal{B} := \mathcal{P}_d$ (suitably normalized), or more generally as $\mathcal{P}_d^n$ when n qudits are to be considered. For a single qudit, we can index the basis elements as $\mathcal{B} = \{P_i\}_{i=1}^{d^2}$. Note that this particular choice of basis has the useful property that it forms a Hermitian, trace orthonormal basis of $B(\mathcal{H})$ so that $P_i^\dagger = P_i$ and $Tr(P_i^\dagger P_j) = \delta_{ij}$, where $\langle A, B\rangle := Tr(A^\dagger B)$ is the Hilbert-Schmidt inner product which endows $D(\mathcal{H})$ with its Hilbert space structure.

Given such a choice we can expand operators on $B(\mathcal{H})$ with respect to this basis. A state $\rho \in D(\mathcal{H})$ can be expressed as

$$\rho = \sum_{i=1}^{d^2} Tr(P_i \rho) P_i. \tag{A.4}$$

Similarly, a CP map $\mathcal{E}$ given in terms of a Kraus decomposition can be expressed as

$$\mathcal{E}(\rho) = \sum_j K_j \rho K_j^\dagger = \sum_{m,n} \chi_{mn} P_m \rho P_n, \tag{A.5}$$

where we have expanded each Kraus operator $K_j \in B(\mathcal{H})$ as

$$K_j = \sum_{m=1}^{d^2} \alpha_{jm} P_m \tag{A.6}$$

and introduced the coefficients

$$\chi_{mn} := \sum_j \alpha_{jm} \alpha_{jn}^*. \tag{A.7}$$

The matrix χ whose coefficients are defined to be $[\chi]_{m,n} := \chi_{mn}$ is the χ-matrix representation of the map $\mathcal{E}$, which completeley determines the action of $\mathcal{E}$ on any quantum state. In general, if $\mathcal{H}$ is a d-dimensional Hilbert space, χ is a $d^2 \times d^2$ matrix, and thus has a total of d^4 paramters. However, due to the CP/TP conditions on the map $\mathcal{E}$ the χ matrix will have $d^2 - d^2$ independent parameters. Since the Hilbert space dimension typically grows exponentially, i.e. as $d = 2^n$ for a n-qubit system, determining all the free parameters of χ is itself exponential.

REFERENCES

[1] I. L. Chuang and M. A. Nielsen *Prescription for experimental determination of the dynamics of a quantum black box (1996)* arXiv:9610001

[2] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing Quantum Gates via Randomized Benchmarking (2012)*, arXiv:1109.6887

[3] E. Magesan, et. al., *Efficient measurement of quantum gate error by interleaved randomized benchmarking (2012)*, arXiv:1203.4550

[4] A. Carignan-Dugas, J. Wallman, J. Emerson, *Characterizing Universal Gate Sets via Dihedral Benchmarking (2015)*, arXiv:1508.06312

[5] A. Cross, et. al., *Scalable randomized benchmarking of non-Clifford gates (2015)*, arXiv:1510.02720

[6] J. Wallman, M. Barnhill, J. Emerson, *Robust Characterization of Loss Rates (2015)*, arXiv:1510.01272

[7] J. Wallman, et. al., *Estimating the Coherence of Noise (2015)*, arXiv:1503.07865

[8] J. Gambetta, et. al., *Characterization of addressability by simultaneous randomized benchmarking (2012)* , arXiv:1204.6308

[9] C. Dankert, et. al., *Exact and Approximate Unitary 2-Designs: Constructions and Applications (2009)* , arXiv:0606161v2

[10] D. Gottesman, *The Heisenberg Representation of Quantum Computers (1998)* , arXiv:9807006v1

[11] J. Wallman, *Randomized benchmarking with gate-dependent noise (2018)* , arXiv:1703.09835v4

[12] J. Wallman, S. Flammia, *Randomized Benchmarking with Confidence (2014)*, arXiv:1404.6025

[13] J. Epstein, et. al., *Investigating the limits of randomized benchmarking protocols (2014)*, arXiv:1308.2928

[14] J. Wallman, J. Emerson, *Noise tailoring for scalable quantum computation via randomized compiling (2016)*, arXiv:1512.01098