Representing and Probing Errors in Quantum Information Processing Devices

by

Junan Lin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2018

Abstract

The quality of quantum information processing devices has been improving at an unprecedented speed. How to faithfully represent the quality of these devices has become an increasingly imminent problem. In this thesis we focus on two aspects in representing and characterizing quantum devices. First, we discuss why most conventional quality metrics are not in principle appropriate to quantify experimentally-determined representations of gate-set elements, due to a gauge degree of freedom in quantum experiments. We then propose an operational quality measure for a gate-set and discuss its usefulness in representing degree of errors and improving experimental control. Second, we develop a protocol that separately and unambiguously characterizes state and measurement errors, relying on high-quality quantum gates. By integrating a method called randomized compiling, we derive a favorable upper bound for the effects of gate errors on the estimated parameters, and numerically demonstrate its performance in the presence of an adversarial gate error.

Acknowledgements

First and foremost I would like to express my deepest gratitude to my supervisor Raymond Laflamme for all his guidance and assistance during my graduate studies. Every discussion with Ray has been both inspiring and enjoyable, from which I gradually learned how research should be performed from a very high level. I'm also thankful towards my advisory committee members Jonathan Baugh and Kevin Resch, who provided valuable suggestions on my research directions as well as my presentation techniques.

I would like to thank my major collaborator and (unofficial) academic advisor Joel Wallman, who has been extremely patient with my countless questions both in and out of research and provided enormous help during my studies. I would also like to thank Brandon Buonacorsi and Nayeli Azucena Rodriguez Briones for both being extremely nice collaborators, with whom we developed many great ideas together. As well, I would like to thank Stefanie Beale, Hemant Katiyar, Michael Chen and Shayan Majidy for all being wonderful groupmates. I would also acknowledge Dawei Lu who has provided extremely helpful guide in research when I first got into the group, as well as being a role model in life.

Special thanks goes to my officemate and good friend Christopher Chamberland, who has inspired me to work hard towards my goal and introduced me to many new sports; and Theerapat Tansuwannont, who has always been a great friend with whom I can discuss anything: thanks for all the great memories which I will carry for long. I would also like to thank Maria Papageorgiou, Keren Li, Hengyan Wang, Jun Li, Zhihuang Luo for all the helpful discussions. I'm grateful to my parents who has always been encouraging throughout my degree. Many thanks to Steve Weiss who provided me with the heavylift machine and introduced me to Linux. As well, I appreciate all the help from the IQC and Physics staff, without whom I would be stuck forever with miscellaneous trivialities in life.

Lastly, many thanks for my examination committee (Professor Raymond Laflamme, Jonathan Baugh, Rajibul Islam, Joel Wallman) for their time reading my thesis. Fundings for this research was provided by NSERC, the Department of Physics and Astronomy, and the Institute for Quantum Computing.

Table of Contents

List of Figures

List of Symbols

$\mathcal{C}(\mathcal{H}^n, \mathcal{H}^m)$ The set of completely positive, trace preserving maps in $T(\mathcal{H}^n, \mathcal{H}^m)$, equivalently known as quantum channels, which will be abbreviated as $\mathcal{C}(\mathcal{H}^n)$ if $n = m$. 6

$E_{a,b}$ Standard basis for $L(\mathcal{H}^m \otimes \mathcal{H}^n)$, which can be represented as a matrix with input 1 on the (a, b)-th position and 0 elsewhere. 8

$\mathbf{P}_n$ The set of n-qubit Pauli operators, normalized with respect to the Hilbert-Schmidt inner product. 9

$T(\mathcal{H}^n, \mathcal{H}^m)$ The set of linear maps from $L(\mathcal{H}^n)$ to $L(\mathcal{H}^m)$, which will be abbreviated as $T(\mathcal{H}^n)$ if $n = m$. 6

$\mathbb{C}^n$ A complex Euclidean space with dimension n. 3

$\mathbb{D}^n$ The set of density operators in $L(\mathcal{H}^n)$. 4

e_i A standard (column) basis vector in $\mathbb{C}^n$. 3

$\mathbf{Herm}(\mathcal{H}^n, \mathcal{H}^m)$ The set of Hermitian operators from $\mathcal{H}^n$ to $\mathcal{H}^m$. 3

$\mathcal{H}^n$ A Hilbert space with dimension n. 3

$\mathbb{1}_{L(\mathcal{H}^d)}$ The identity channel in $\mathcal{H}^d$. 6

$\mathbb{1}_{\mathcal{H}^d}$ The identity operator in $\mathcal{H}^d$. 5

$|\psi\rangle$ A column vector ψ written in Dirac notation. 3

$L(\mathcal{H}^n, \mathcal{H}^m)$ The set of linear operators that takes vectors in $\mathcal{H}^n$ to $\mathcal{H}^m$. If the input and output spaces have the same dimension n, it will be abbreviated as $L(\mathcal{H}^n)$. 3

$\boldsymbol{Pos}(\mathcal{H}^n, \mathcal{H}^m)$ The set of positive-semidefinite operators from $\mathcal{H}^n$ to $\mathcal{H}^m$. 4

$\mathcal{U}(\mathcal{H}^n)$ The set of unitary channels in $T(\mathcal{H}^n)$. 8

$U(\mathcal{H}^n)$ The set of unitary operators in $L(\mathcal{H}^n)$. 8

Chapter 1

Introduction

Quantum computation is a novel computational approach that utilizes the laws of quantum mechanics [1]. It is capable of solving certain important problems much faster compared to all currently known classical algorithms [2, 3, 4]. The task of building a reliable quantum computer has become an increasingly important task in the past two decades.

Quantum computation requires a different architecture compared to classical computers. In the circuit model of quantum computing, a computational task is divided into preparing a certain state, performing particular quantum gates to transform the state, and making a final measurement of the resultant state. Since the degree to which we can perform these operations with high fidelity is the key to achieving successful computation, every one of these steps contributes to the final success rate. In order to improve experimental control, it is crucial to be able to faithfully detect errors in a quantum information processing (QIP) device. The word faithful is key to our discussion here: first, if an error-identifying protocol incorrectly assigns the non-ideal behavior of a device to a wrong component, it can cause unnecessary confusion when one tries to make corrections to lower the error rates. For example, one could waste both time and money trying to improve the quality of a single photon source for months, before finally finding out that it is actually the miscalibrated detector that gives undesired results. Second, if the protocol underestimates the severity of errors, it may result in unexpected failures in computations. Up to today, quantum gates that are claimed to be close to or surpassing the fault-tolerance threshold for certain error correcting codes have been reported in many platforms including superconducting qubits [5], quantum dots [6, 7], trapped ions [8], and solid state nuclear spins [9, 10]. However, it has also been pointed out that there may be mismatch between the claimed numbers and the truly desired theoretical figures of merit [11].

The focus of this thesis will be on methods of representing and detecting errors in a QIP device. The main contribution of this thesis is as follows: first, we address the problem of whether any gauge-invariant quality metric reported in a tomography experiment can be fully trusted, with a negative answer. We also propose an alternative distance measure for a gate-set which depends on measurable quantities only, and discuss its usefulness in characterizing and calibrating QIP devices. These aspects are discussed in Chapter 4. Second, we provide a protocol to separately characterize state and measurement errors, and derive bounds on the parameters which can be estimated independently of their errors. We show its performance by simulating it numerically in the presence of an adversarial gate error, demonstrating its experimental reliability. These aspects are discussed in Chapter 5. In addition, in Chapter 2 we define the necessary mathematical concepts that are relevant in quantum information. A short review on different protocols that aim to characterize a quantum device is given in Chapter 3.

Chapter 2

Mathematical Background

We introduce some important mathematical concepts that are used in this work. Notations used here largely follow the ones in [12], which is also the main source, although multiple adaptations exist. This section also references the "Mathematical Background" chapter in the thesis by Puzzuoli [13] and the "Introduction" chapter by Magesan [14].

2.1 States

Quantum mechanics allows us to describe states of objects using vectors in a Hilbert space $\mathcal{H}^n$ with dimension n, which is isomorphic to the complex Euclidean space $\mathbb{C}^n$ equipped with the Euclidean inner product $\langle \cdot \rangle$. Defining a standard basis e_i for $\mathbb{C}^n$ which is a column vector having entry 1 at position i and 0 elsewhere, it is easily seen that $\langle e_i, e_j \rangle = \delta_{ij}$, and any vector can be defined in terms of its components in this basis. Under Dirac's notation, the vector is typically represented as $|\psi\rangle$; in component form, this can be written as $|\psi\rangle = \psi_i e_i$ where the summation is implicit through the Einstein convention. For every such vector one can define its conjugate transpose denoted as $\langle\psi|$: specifically, $\langle\psi| = \psi_i^* e_i^T$ where the superscript * denotes complex conjugate. The inner product is given by $\langle \psi, \phi \rangle = \psi_i^* \phi_i$, which is represented in Dirac notation as $\langle\psi|\phi\rangle$. In quantum mechanics, this number is regarded as the *overlap* between the states represented by $|\psi\rangle$ and $|\phi\rangle$.

The outer product between two vectors $|\phi\rangle \in \mathcal{H}^m$ and $\langle\psi| \in \mathcal{H}^n$ is given by $|\phi\rangle\langle\psi|$. This forms a (linear) *operator* in the space we call $L(\mathcal{H}^n, \mathcal{H}^m)$, which takes vectors in $\mathcal{H}^n$ to ones in $\mathcal{H}^m$. We will often shorthand $L(\mathcal{H}^n, \mathcal{H}^n)$ as $L(\mathcal{H}^n)$. An important class of operators in $L(\mathcal{H}^n)$ are *Hermitian operators*, denoted as $\mathrm{Herm}(\mathcal{H}^n, \mathcal{H}^m)$: an operator A is

Hermitian if $A = A^\dagger$ where † denotes the conjugate transpose of a matrix. This family of operators are of particular interest because it is a postulate in quantum mechanics that all physical observables are represented as Hermitian operators. Note also that Hermitian operators have real eigenvalues, which enables us to discuss its spectrum being positive or negative. Among Hermitian operators, those with non-negative eigenvalues are called *positive-semidefinite* operators, denoted as $Pos(\mathcal{H}^n, \mathcal{H}^m)$.

The space of operators $L(\mathcal{H}^n, \mathcal{H}^m)$ can be made into an inner product space with the following definition:

Definition 2.1.1. The *Hilbert-Schmidt inner product* between two operators $A, B \in L(\mathcal{H}^n, \mathcal{H}^m)$ is defined as

$$\langle A, B\rangle_{HS} \equiv \langle A, B\rangle := \mathrm{Tr}\big[A^\dagger B\big] \tag{2.1}$$

A quantum system can be composed of different subsystems. A composite system lives in the *tensor product space* of sub-Hilbert spaces, denoted as $\bigotimes_i \mathcal{H}_i$, with a dimensionality equal to the product of that of all subsystems. If two Hilbert spaces $\mathcal{H}^n$ and $\mathcal{H}^m$ have bases vectors $\{|\phi\rangle_i\}$ and $\{|\psi\rangle_j\}$, then the tensor product space $\mathcal{H}^{n+m} = \mathcal{H}^n \otimes \mathcal{H}^m$ has bases $\{|\phi\rangle_i \otimes |\psi\rangle_j\}$.

2.1.1 Density operators

In addition to *pure states* described by a single state vector, one may also consider a statistical ensemble of quantum systems, each in a different pure state: in such cases the state of a single system in this ensemble cannot be represented by any particular state vector anymore. Such states are called *mixed states* and can be described by a type of Hermitian operator called *density operators*, defined as follows:

Definition 2.1.2. $\rho \in L(\mathcal{H}^n)$ is called a density operator if

$$\rho \in \mathrm{Pos}(\mathcal{H}^n),\ \ \mathrm{Tr}(\rho) = 1$$

We denote the set of density operators in $\mathrm{Herm}(\mathcal{H}^n)$ by $\mathbb{D}^n$.

Every $\rho \in \mathbb{D}^n$ can be written as the following linear combination

$$\rho = \sum_i p_i \, |\psi_i\rangle \, \langle\psi_i| \tag{2.2}$$

where $|\psi_i\rangle$ forms a complete basis for $\mathcal{H}^n$, and the p_i's correspond to the weights in the statistical ensemble. In particular, for every pure state $|\phi\rangle \in \mathcal{H}^n$, there is a unique density operator $\rho = |\phi\rangle \langle\phi|$, which is a rank-1 projector in the space $L(\mathcal{H}^n)$. One may define the *purity* of ρ as $\mathrm{Tr}(\rho^2)$, based on the observation that this number is 1 only when ρ corresponds to a pure state, and is less than 1 if it is a mixed state.

The most fundamental computational units in quantum information science are two-level quantum systems, or *qubits*. The state of a qubit can be described by a density operator $\rho \in \mathrm{Herm}(\mathcal{H}^2)$. A particularly well-known basis for $\mathrm{Herm}(\mathcal{H}^2)$ is the set of Pauli matrices including the identity matrix, which has the elements

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The above set is often referred to as the *single-qubit Pauli matrices*, denoted as $\mathbf{P}_1$. It can be easily verified that this basis is orthogonal with respect to the Hilbert-Schmidt inner product, and can become an orthonormal basis if each multiplied by a normalization factor.

2.2 Measurement

Measurements play an important role in quantum mechanics since all quantum phenomena are detected through some form of measurement. Unlike classical measurements, quantum measurements can potentially change the state of the system being measured, and non-commuting measurements have precision limited by the uncertainty principle. The most general formalism for quantum measurements is the positive operator-valued measure (POVM) formalism: mathematically, a measurement $\mathcal{M}$ is described in terms of a set of measurement operators $M_i \in \mathrm{Pos}(\mathcal{H}^d)$, satisfying

$$\sum_i M_i = \mathbb{1}_{\mathcal{H}^n} \tag{2.3}$$

where $\mathbb{1}_{\mathcal{H}^d}$ denotes the identity operator in $\mathcal{H}^d$. These operators are indexed by their corresponding *outcomes* i. For a system originally in the state represented by $\rho \in \mathbb{D}(\mathcal{H}^d)$, *Born's rule* states that the probability for obtaining the outcome i is given by

$$p_i = \mathrm{Tr}[M_i \rho] \tag{2.4}$$

This defines a valid probability distribution because $p_i \geq 0 \; \forall \; i$, and $\sum_i p_i = 1$: the second condition holds simply because $\rho \in \mathbb{D}^n$, and

$$\sum_i p_i = \mathrm{Tr}\left[\sum_i M_i \rho\right] = \mathrm{Tr}[\rho] = 1 \tag{2.5}$$

To show that the first condition holds, we decompose the Hermitian operators M_i and ρ using the spectral theorem:

$$\begin{aligned} M_i &= \sum_k m_{i,k} \left|\psi_k\right\rangle \left\langle\psi_k\right| \\ \rho &= \sum_j r_j \left|\phi_j\right\rangle \left\langle\phi_j\right| \end{aligned} \tag{2.6}$$

where $0 \leq m_{i,k} \leq 1$ and $0 \leq r_j \leq 1$ are the eigenvalues of M_i and ρ. Therefore

$$\begin{aligned} \mathrm{Tr}[M_i \rho] &= \mathrm{Tr}\left[\sum_{j,k} r_j m_{i,k} \left|\phi_j\right\rangle \left\langle\phi_j|\psi_k\right\rangle \left\langle\psi_k\right|\right] \\ &= \sum_{j,k} r_j m_{i,k} |\langle\phi_j|\psi_k\rangle|^2 \geq 0 \end{aligned} \tag{2.7}$$

since every term in the sum is non-negative.

2.3 Evolution

A quantum system can evolve from one state to another. This evolution may be described by a linear map from $L(\mathcal{H}^n)$ to $L(\mathcal{H}^m)$, a set which is denoted as $T(\mathcal{H}^n, \mathcal{H}^m)$. In particular, we will shorthand the set $T(\mathcal{H}^n, \mathcal{H}^n)$ as $T(\mathcal{H}^n)$. For a process to be physically allowed, we require that a map $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$ satisfies the following properties: for all $A \in \mathcal{H}^{n+k}$ and $A \geq 0$,

$$(\Phi \otimes \mathbb{1}_{L(\mathcal{H}^k)})(A) \geq 0 \; \forall k \in \mathbb{Z}^+, \; \mathrm{Tr}(A) = \mathrm{Tr}(\Phi(A)) \tag{2.8}$$

where $\mathbb{1}_{L(\mathcal{H}^d)}$ denotes the identity map from $\mathcal{H}^d$ to $\mathcal{H}^d$. We call any Φ that satisfies the above property a completely-positive, trace-preserving (CPTP) map, also called a quantum *channel* whose set is denoted as $\mathcal{C}(\mathcal{H}^n, \mathcal{H}^m)$. This definition is based on realistic consideration that any physically allowed process acting on a valid state must return a valid state, and this must hold whether we consider the state on its own or being part of some larger system. A CPTP map is the most general description of an evolution happening on a quantum system with no initial correlation with the environment [14, 1]. Naturally, maps that satisfy only the first condition are called completely-positive (CP), and the second, trace-preserving (TP). We summarize the *canonical constraints* regarding representations for quantum objects: that states are represented by density operators, measurements by measurement operators, and gates by CPTP maps.

Given a map $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$, there exists a unique *adjoint* map $\Phi^\dagger \in T(\mathcal{H}^m, \mathcal{H}^n)$, defined through the following equation:

$$\langle A, \Phi(B) \rangle = \langle \Phi^\dagger(A), B \rangle \ \forall A \in L(\mathcal{H}^m),\ B \in L(\mathcal{H}^n) \tag{2.9}$$

Next, we introduce some of the most common representations for quantum maps that are relevant to this thesis.

2.3.1 Kraus representation

The effect of every linear map $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$ can be written as

$$\Phi(\rho) = \sum_i A_i \rho B_i^\dagger \tag{2.10}$$

with some $A_i,\ B_i \in L(\mathcal{H}^n, \mathcal{H}^m)$. This is called the *Kraus representation* of the map Φ and A_i, B_i are called the Kraus operators. For every map Φ represented as in 2.10, one can find its adjoint map as

$$\Phi^\dagger(\sigma) = \sum_i A_i^\dagger \sigma B_i \tag{2.11}$$

This can be derived by directly applying the definition of an adjoint map, Equation 2.9:

$$\begin{aligned} \langle X, \Phi(Y) \rangle &= \mathrm{Tr}\left[\sum_i X^\dagger A_i Y B_i^\dagger\right] = \sum_i \mathrm{Tr}\left[B_i^\dagger X^\dagger A_i Y\right] \\ &= \mathrm{Tr}\left[\sum_i (A_i^\dagger X B_i)^\dagger Y\right] \\ &\equiv \langle \Phi^\dagger(X), Y \rangle \end{aligned} \tag{2.12}$$

where we made use of the facts that the order of summation and trace can be exchanged, and the trace has a cyclic property, $\mathrm{Tr}[ABC] = \mathrm{Tr}[CAB] = \mathrm{Tr}[BCA]$.

One advantage of writing a map in the Kraus representation is that complete positivity can be trivially verified given the set of Kraus operators. It is known [12] that a map Φ is CP iff there exists $A_i \in T(\mathcal{H}^n, \mathcal{H}^m)$ such that

$$\Phi(\rho) = \sum_i A_i \rho A_i^\dagger \tag{2.13}$$

and is a channel if, in addition to 2.13, the following also holds

$$\sum_i A_i^\dagger A_i = \mathbb{1}_{\mathcal{H}^n} \tag{2.14}$$

This means that a map being CP is equivalent to having $A_i = B_i \ \forall i$. This property is summarized later in Theorem 2.3.1.

An important class of channels are *Unitary channels*, which can be represented with one unitary Kraus operator:

Definition 2.3.1. An operator $V \in L(\mathcal{H}^n)$ is called a *unitary operator* if it satisfies $VV^\dagger = V^\dagger V = \mathbb{1}$. The set of unitary operators in $L(\mathcal{H}^n)$ is denoted as $U(\mathcal{H}^n)$.

Definition 2.3.2. A channel $\Phi \in \mathcal{C}(\mathcal{H}^n)$ is called a unitary channel if $\Phi(A) = VAV^\dagger \ \forall A \in L(\mathcal{H}^n)$, for some $V \in U(\mathcal{H}^n)$. The set of unitary channels acting on operators in $L(\mathcal{H}^n)$ is denoted as $\mathcal{U}(\mathcal{H}^n)$.

2.3.2 Choi representation

The Choi representation provides an alternative way to look at quantum maps [12].

Definition 2.3.3. The Choi representation for a map $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$ is a mapping $J : T(\mathcal{H}^n, \mathcal{H}^m) \to L(\mathcal{H}^m \otimes \mathcal{H}^n)$ defined as

$$J(\Phi) = \sum_{a,b} \Phi(E_{a,b}) \otimes E_{a,b} \tag{2.15}$$

where $E_{a,b}$ are the standard basis for $L(\mathcal{H}^m \otimes \mathcal{H}^n)$, which can be represented as a matrix with input 1 on the (a, b)-th position and 0 elsewhere.

The Choi representation is unique for each map as Equation 2.15 is a linear bijection. It offers a convenient way to determine whether a map is CPTP, as established by the theorem below [15, 12].

Theorem 2.3.1. *Let $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$ be a non-zero map. The following statements are equivalent:*

1. *Φ is CPTP.*
2. *$J(\Phi) \in Pos(\mathcal{H}^m \otimes \mathcal{H}^n)$ and $\mathrm{Tr}_2[J(\Phi)] = \mathbb{1}_{\mathcal{H}^n}$.*
3. *There exists $\{A_i\} \subset L(\mathcal{H}^n, \mathcal{H}^m)$ such that $\Phi(X) = \sum_i A_i X A_i^\dagger \ \forall X \in L(\mathcal{H}^n)$, and $\sum_i A_i^\dagger A_i = \mathbb{1}_{\mathcal{H}^n}$.*

where Tr_2 is the partial trace operation defined as

$$\mathrm{Tr}_2[A \otimes B] = \mathrm{Tr}[B]A \ \forall \ A \in L(\mathcal{H}^n), B \in L(\mathcal{H}^m) \tag{2.16}$$

The above theorem says that for every CPTP map in $T(\mathcal{H}^n, \mathcal{H}^m)$ uniquely corresponds to a positive-semidefinite operator in $L(\mathcal{H}^m \otimes \mathcal{H}^n)$. It also summarizes the relationship between Kraus operators and the CPTP property of Φ.

2.3.3 Liouville representation

The third characterization that we present involves both operators and quantum maps. Similar to the Choi representation, the Liouville representation also express the map as a matrix (a “superoperator” that acts on operators), but using a different mapping that allows straight-forward calculation of how states transform under the map. Here we present the slightly less general case when the states are composed of qubits, which is the most relevant situation in quantum information[16]. We have seen that the density matrix of a single qubit can be expressed as a real linear combination of elements in $\mathbf{P_1}$. For a system of n-qubits, one can similarly define $\mathbf{P}_n$ to be an orthonormal basis for $\mathrm{Herm}(\mathcal{H}^{2^n})$, where

$$\mathbf{P}_n = \left\{ P_k : P_k = \bigotimes_x \left(\frac{\sigma_{x_k}}{\sqrt{2}} \right) \right\}$$

where x is an n-tuple generated from the list $\{x, y, z, I\}$, and k goes from 1 to 4^n. It has the same orthonormal property as $\mathbf{P}_1$:

$$\langle P_i,\ P_j \rangle_{HS} \coloneqq \mathrm{Tr}\left[P_i^\dagger P_j \right] = \delta_{ij}$$

This basis allows us to "vectorize" density matrices for systems of qubits. Specifically, any 2^n by 2^n Hermitian matrix can be decomposed in terms of their relative inner products with respect to the Pauli basis. Writing these as components of a vector will define a representation in the space of $2^{2n} \times 1$ real vectors, which is isomorphic to the set of $2^n \times 2^n$ real matrices as can be seen from the vec map between the basis vectors between the two:

$$\mathrm{vec}(E_{a,b}) = e_{2^n \times b + a} \tag{2.17}$$

where $E_{a,b}$ forms the $(2^n \times 2^n)$ standard (operator) basis, and e_i forms the $(2^{2n} \times 1)$ standard (vector) basis (having 1 at the i-th component and 0 elsewhere)[1]. In other words, the vec map is simply taking the columns of a matrix and stacking them to form a vector.

For every 2^n by 2^n Hermitian matrix ρ, we define its Liouville representation as:

$$|\rho\rangle\rangle := \sum_i \mathrm{Tr}[\rho P_i] |i\rangle\rangle \tag{2.18}$$

where $|i\rangle\rangle \equiv e_i$ denotes the standard vector basis in this representation. One can define an element σ in the dual space (representing a measurement operator) as

$$\langle\langle\sigma| := \sum_i \mathrm{Tr}[P_i \sigma]^* \langle\langle i| \tag{2.19}$$

where * denotes complex conjugation. We see that the Hilbert-Schmidt inner product is now transformed into an Euclidean inner product:

$$\begin{aligned}
\langle\langle\sigma|\rho\rangle\rangle &= \sum_{i,j} \mathrm{Tr}[\rho P_i]\, \mathrm{Tr}[P_j \sigma]^* \langle\langle j|i\rangle\rangle \\
&= \sum_{i,j} \mathrm{Tr}[\rho P_i]\, \mathrm{Tr}[P_j \sigma]^*\, \mathrm{Tr}[P_i P_j] \\
&= \mathrm{Tr}\left[\sum_i \mathrm{Tr}[\rho P_i] P_i \sum_j \mathrm{Tr}[P_j \sigma]^* P_j\right] \\
&= \mathrm{Tr}\left[\sigma^\dagger \rho\right]
\end{aligned} \tag{2.20}$$

Next, we define how maps can be represented in this basis. Since $\mathbf{P_n}$ forms a complete basis for $L(\mathcal{H}^{2^n})$, specifying how the map acts on every element of $\mathbf{P_n}$ is sufficient to quantify its exact behavior because the map Φ is linear by the axioms of quantum mechanics.

---

[1] Note that here a and b go up to 2^n whereas i and j go up to 2^{2n}.

Definition 2.3.4. The *Liouville representation* (also called the *Pauli Transfer Matrix, PTM, in some literature*) for a map $\Phi \in T(\mathcal{H}^{2^n})$ is a $(4^n \times 4^n)$ matrix G_Φ with components

$$(G_\Phi)_{i,j} := \mathrm{Tr}[P_i \Phi(P_j)] \tag{2.21}$$

Using this representation, the action of Φ on a state ρ can be straightforwardly calculated as a matrix-vector multiplication:

$$\begin{aligned} |\Phi(\rho)\rangle\rangle &= \sum_i \mathrm{Tr}[\Phi(\rho)P_i]|i\rangle\rangle \\ &= \sum_i \mathrm{Tr}\left[\Phi(\sum_j \mathrm{Tr}[\rho P_j]P_j)P_i\right]|i\rangle\rangle \\ &= \sum_{ij} \mathrm{Tr}[\rho P_j]\,\mathrm{Tr}[\Phi(P_j)P_i]|i\rangle\rangle \\ &= \sum_i \sum_j (G_\Phi)_{i,j}(|\rho\rangle\rangle)_j|i\rangle\rangle \\ &= G_\Phi|\rho\rangle\rangle \end{aligned} \tag{2.22}$$

due to linearity of Φ.

Another advantage of this representation is that the TP property for a map is very easy to determine[16]: since $\mathrm{Tr}[P_i] = \delta_{0,i}$, a TP map must satisfy $\mathrm{Tr}[\Phi(P_i)] = \delta_{0,i}$. On the other hand,

$$\mathrm{Tr}[\Phi(P_i)] = \mathrm{Tr}[P_0\Phi(P_i)] = (G_\Phi)_{0,i} \tag{2.23}$$

so Φ is TP iff $(G_\Phi)_{0,i} = \delta_{0,i}$: i.e., the first row of G_Φ is a 1 followed by all 0's. However, the CP property cannot be easily determined directly from this representation. Fortunately, since both the Liouville and Choi representations are bijective mappings from the actual map, it is possible to determine the TP condition using Equation 2.23, and the CP condition from Theorem 2.3.1. A formula that converts between these two representations is derived in the Appendix A.

2.4 Quantum Information Basics

2.4.1 Bloch Sphere and Pauli Observables

As mentioned before, the basic computational units in a quantum computer are two-level quantum systems called *qubits*, in analogy to bits in a classical computer. One typically denotes the state space of a qubit as spanned by the *computational basis vectors* $\{|0\rangle, |1\rangle\}$. A particularly intuitive way to represent a single-qubit state is by using a Bloch-sphere (Figure 2.1), by noting that the state vector of any 2-level quantum system can be written with the following parametrization,

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle\,,\ 0 \leq \theta \leq \pi,\ 0 \leq \phi \leq 2\pi \tag{2.24}$$

where we make use of the fact that relative phases between basis states are not physically relevant, and that a pure state satisfies the completeness condition $\langle\psi|\psi\rangle = 1$. The density operator corresponding to this general pure state is

$$\rho_\psi = |\psi\rangle\langle\psi| = \begin{pmatrix} \cos^2\left(\frac{\theta}{2}\right) & e^{-i\phi}\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right) \\ e^{i\phi}\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right) & \sin^2\left(\frac{\theta}{2}\right) \end{pmatrix} \tag{2.25}$$

which can be written in the Liouville representation as

$$|\rho\rangle\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \vec{u} \end{pmatrix} \tag{2.26}$$

where we identified the vector $\vec{u}$ as a unit vector in $\mathbb{R}^3$ with the parameters θ and ϕ being the spherical coordinates on a unit sphere. This allows a convenient visualization for both the states and their transformations. We see that pure states have 1-to-1 correspondence to points on the surface of the sphere, and among these, computational basis vectors are shown as north and south antipodal points. Mixed states, on the other hand, can be written as convex combinations of pure states, and thus correspond to points in the interior of the Bloch sphere.

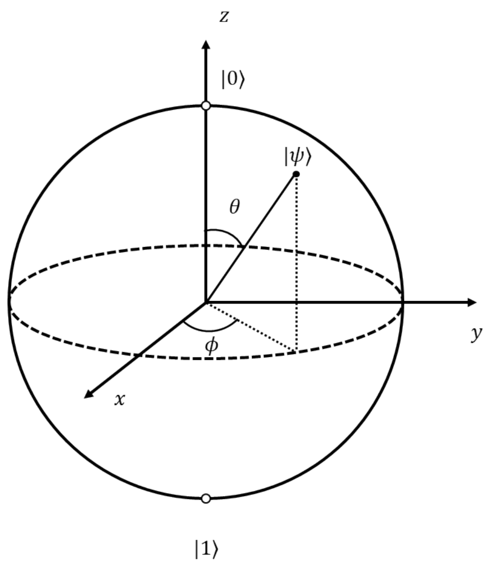

Figure 2.1: Representation of a pure qubit state on a Bloch sphere.

2.4.2 Quantum circuit diagrams

In analogy to using a circuit diagram to represent how a classical computational algorithm is carried out by showing the operations being applied at different time steps, a quantum circuit diagram shows what transformation is being performed to the system in chronological order (usually from left to right), with horizontal lines representing different qubits and blocks representing quantum gates. Figure 2.2 shows an example of such a representation where two states $|\rho_1\rangle$ and $|\rho_2\rangle$ are being inputted, a map (also called a quantum gate) $\mathcal{A}$ is applied to the first qubit and $\mathcal{B}$ to the second qubit, then a 2-qubit map $\mathcal{U}$ is applied to both qubits, and qubit 1 is being measured by the measurement $\mathcal{M}$.

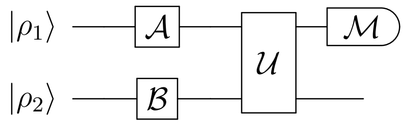

Figure 2.2: Example of a quantum circuit.

Mathematically, the output of this circuit (right before the measurement is applied) can be written as

$$\mathrm{Tr}_2[\mathcal{U}((\mathcal{A}\otimes\mathcal{B})(\rho_i\otimes\rho_2))] \tag{2.27}$$

where Tr_i denotes the action of "tracing out the i-th register". Its mathematical definition is given as follows:

Definition 2.4.1. The *partial trace* map Tr_i is the unique map that satisfies the equation

$$\mathrm{Tr}_i[A_1 \otimes \cdots \otimes A_n] = \mathrm{Tr}[A_i] A_1 \otimes \cdots \otimes A_{i-1} \otimes A_{i+1} \otimes \cdots \otimes A_n \tag{2.28}$$

for all operators $A_1 \ldots A_n$ belonging to potentially different Hilbert spaces.

2.5 Norms and Distance Measures

In quantum information theory it is natural to consider similarity measures between two quantum states or operations: an experimentalist may be interested in how well a quantum computer is performing by examining how close a particular operation is to the ideal. To quantify this similarity, we need to introduce norms and distance measures to the corresponding vector spaces. We define some important examples of such measures that are relevant to this thesis, first for states/measurements (operators), then for quantum gates (maps).

2.5.1 Measures for operators

A vector space V such as $L(\mathcal{H}^n, \mathcal{H}^m)$ can be made into a normed vector space if equipped with a norm, which is a function that assigns "size" to vectors in V.

Definition 2.5.1. A *norm* on $L(\mathcal{H}^n, \mathcal{H}^m)$ is a function $\|\cdot\|$ satisfying the following properties [12]:

1. $\|A\| \geq 0 \ \forall A \in L(\mathcal{H}^n, \mathcal{H}^m)$, with $\|A\| = 0$ iff $A = 0$.
2. $\|\alpha A\| = |\alpha| \|A\| \ \forall A \in L(\mathcal{H}^n, \mathcal{H}^m),\ \alpha \in \mathbb{C}$.
3. $\|A + B\| \geq \|A\| + \|B\| \ \forall A, B \in L(\mathcal{H}^n, \mathcal{H}^m)$ (triangle inequality)

A particularly useful family of norms for operators are the *Schatten p-norms*, defined as follows.

Definition 2.5.2. The Schatten p-norm for an operator $A \in L(\mathcal{H}^n, \mathcal{H}^m)$ and any real number $p \geq 1$ is defined as

$$\|A\|_p := \Big(\mathrm{Tr}\Big((A^\dagger A)^{\frac{p}{2}}\Big)\Big)^{\frac{1}{p}} \tag{2.29}$$

Several properties of the Schatten p-norms are summarized as follows [12]:

1. The Schatten p-norms are non-increasing in p: for every operator $A \in L(\mathcal{H}^n, \mathcal{H}^m)$ and any $1 \leq p \leq q \leq \infty$, it holds that

$$\|A\|_p \geq \|A\|_q \tag{2.30}$$

2. For every $p \in [1, \infty]$, the Schatten p-norm is unitarily invariant: for every $A \in L(\mathcal{H}^n),\ U, V \in \mathcal{U}(\mathcal{H}^n)$ it holds that

$$\|A\|_p = \left\|UAV^\dagger\right\|_p \tag{2.31}$$

3. For every $p \in [1, \infty]$, one may define $p^* \in [1, \infty]$ by

$$\frac{1}{p} + \frac{1}{p^*} = 1 \tag{2.32}$$

   Then for every $A \in L(\mathcal{H}^n, \mathcal{H}^m)$, it holds that the Schatten p-norm and p^*-norm are dual in the sense that

$$\|A\|_p = \max\{|\langle B, A\rangle|\} : B \in L(\mathcal{H}^n),\ \|B\|_{p^*} \leq 1 \tag{2.33}$$

   In particular, this implies the *Hölder's inequality* for Schatten p-norms:

$$|\langle B, A\rangle| \leq \|A\|_p \|B\|_{p^*} \tag{2.34}$$

4. The Schatten p-norm is *submultiplicative*:

$$\|AB\|_p \leq \|A\|_p \|B\|_p \tag{2.35}$$

Among the p-norms, the 1-norm is a particularly popular norm which will also be used in defining other distance measures later. We include one additional property of the 1-norm below [12]:

Lemma 2.5.1. *The Schatten 1-norm is non-increasing under partial tracing: for every operator $A \in L(\mathcal{H}^n \otimes \mathcal{H}^m)$,*

$$\|\mathrm{Tr}_2[A]\|_1 \leq \|A\|_1 \tag{2.36}$$

Next, in order to define a "distance" between objects in the same set, a metric is needed:

Definition 2.5.3. A metric on a set X is a function $d: (X \times X) \to \mathbb{R}$ that satisfies the following requirements:

1. $d(A, B) \geq 0 \; \forall \; A, B \in X$;
2. $d(A, B) = 0$ iff $A = B$;
3. $d(A, B) = d(B, A) \; \forall A, B \in X$;
4. $d(A, C) \leq d(A, B) + d(B, C) \; \forall A, B, C \in X$ (triangle inequality)

Since $L(\mathcal{H}^n, \mathcal{H}^m)$ equipped with any p-norm becomes normed vector space, a metric can be defined on this space based on the norm:

Definition 2.5.4. The *p-norm distance* between two operators $A, B \in L(\mathcal{H}^n, \mathcal{H}^m)$ is defined as

$$d_p(A, B) := \|A - B\|_p \tag{2.37}$$

and the metric is said to be induced by the norm.

While each p-norm distance gives a slightly different measure between two operators, the 1-norm distance is a relatively popular one due to the following theorem, which gives the metric an operational meaning.

Theorem 2.5.2. *(Holevo-Helstrom) Let $\rho_0, \rho_1 \in \mathbb{D}(\mathcal{H}^n)$, and let $\lambda \in [0, 1]$. For every choice of a two-outcome measurement $\mathcal{M} = \{M_0, M_1\}$, it holds that [12]*

$$\lambda \langle M_0, \rho_0 \rangle + (1 - \lambda) \langle M_1, \rho_1 \rangle \leq \frac{1}{2}(1 + \|\lambda \rho_0 - (1 - \lambda \rho_1)\|_1) \tag{2.38}$$

Moreover, there exist a projective measurement $\mathcal{M}$ for which the equality holds.

The LHS of Equation 2.38 is the probability to correctly identify an unknown state between ρ_0 and ρ_1, each given with probability λ and $1-\lambda$, with a single-shot measurement; the RHS is related to the 1-norm distance between these two probabilistic states. This theorem provides a way to calculate the optimal probability given the states, and will become relevant when we define norms and metrics for quantum channels.

In addition to the p-norm distance, another commonly used similarity measure between two positive-semidefinite operators is the fidelity function, whose definition involves the operator 1-norm:

Definition 2.5.5. The *fidelity* between two operators $A, B \in \mathrm{Pos}(\mathcal{H}^n)$ is defined as

$$F(A,B) \coloneqq \left\| \sqrt{A}\sqrt{B} \right\|_1 \tag{2.39}$$

where $\sqrt{P}$ is the unique positive-semidefinite operator that satisfies $\sqrt{P}\sqrt{P} = P$ for $P \in L(\mathcal{H}^n)$.

The fidelity is not a metric: for example, $F(A,A) = 1$ instead of 0. However, there are some properties of this definition that makes it useful under many situations: in particular, let $A, B \in \mathrm{Pos}(\mathcal{H}^n)$ be positive semidefinite operators. Then the following facts hold [12]:

1. $F(A,B) = F(B,A)$;
2. $F(\lambda A, B) = F(A, \lambda B) = \sqrt{\lambda} F(A,B)\ \forall \lambda \in \mathbb{R}^+$;
3. $F(A,B) \geq 0$ with equality iff $AB = 0$;
4. Unitary invariance: for every unitary operator $V \in \mathcal{U}(\mathcal{H}^n)$, it holds that $F(A,B) = F(VAV^\dagger, VBV^\dagger)$.

Using the definition of the 1-norm in Equation 2.29, an alternative expression for the fidelity function can be derived as

$$\begin{aligned} F(A,B) &= \mathrm{Tr}\left[\sqrt{\sqrt{B}^\dagger \sqrt{A}^\dagger \sqrt{A}\sqrt{B}}\right] \\ &= \mathrm{Tr}\left[\sqrt{\sqrt{B} A \sqrt{B}}\right] \end{aligned} \tag{2.40}$$

which holds since both $\sqrt{A}$ and $\sqrt{B}$ are positive-semidefinite. In the case where either A or B represents a pure state, the above can be further simplified to

$$\begin{aligned} F(A,B) &= \mathrm{Tr}\left[\sqrt{\sqrt{|\psi\rangle\langle\psi|} A \sqrt{|\psi\rangle\langle\psi|}}\right] \\ &= \sqrt{\langle\psi| A |\psi\rangle}\, \mathrm{Tr}\left[\sqrt{|\psi\rangle\langle\psi|}\right] \\ &= \sqrt{\langle\psi| A |\psi\rangle} \end{aligned} \tag{2.41}$$

where WLOG we assumed that B is a pure state, $B = |\psi\rangle\langle\psi|$.

2.5.2 Measures for quantum maps

Deriving from distance measures for quantum states represented by operators, one may define distance measure for quantum maps to characterize the closeness between two operations. In general, these measures can be defined as either an averaged or optimized value of the corresponding operator distance over some set [17]. The two most commonly quoted measures are given as follows: the first one is induced from the fidelity (or more precisely, the square of the fidelity as defined in Equation 2.39),

Definition 2.5.6. The *average gate infidelity*, r, between a TP map $\mathcal{E} \in T(\mathcal{H}^n)$ and a unitary channel $\mathcal{V} \in \mathcal{U}(\mathcal{H}^n)$ is defined as [1, 18]

$$r(\mathcal{E},\mathcal{V}) \coloneqq 1 - \int d\psi \;\; \mathrm{Tr}[\mathcal{E}(|\psi\rangle\langle\psi|)\mathcal{V}(|\psi\rangle\langle\psi|)] \equiv 1 - \bar{F}(\mathcal{E},\mathcal{V}) \tag{2.42}$$

where the integral is taken over the unitarily-invariant Haar measure[2]. Alternatively, one may also define equivalently a single argument version as

$$r(\mathcal{E},\mathcal{V}) = r(\mathcal{V}^\dagger\mathcal{E},\mathbb{1}) \coloneqq r(\mathcal{V}^\dagger\mathcal{E}) = 1 - \int d\psi \;\; \langle\psi|\,\mathcal{V}^\dagger\mathcal{E}(|\psi\rangle\langle\psi|)\,|\psi\rangle \equiv 1 - \bar{F}(\mathcal{V}^\dagger\mathcal{E}) \tag{2.43}$$

The first equality in Equation 2.43 can be seen using the Kraus representation of $\mathcal{V}$: since $\mathcal{V}$ is a unitary channel,

$$\mathcal{V}(X) = UXU^\dagger \tag{2.44}$$

for some unitary operator U. Thus,

$$\begin{aligned} r(\mathcal{V}^\dagger\mathcal{E},\mathbb{1}) &= 1 - \int d\psi \;\; \mathrm{Tr}\big[\mathcal{V}^\dagger\mathcal{E}(|\psi\rangle\langle\psi|)(|\psi\rangle\langle\psi|)\big] \\ &= 1 - \int d\psi \;\; \mathrm{Tr}\left[U^\dagger(\mathcal{E}(|\psi\rangle\langle\psi|))U(|\psi\rangle\langle\psi|)\right] \\ &= 1 - \int d\psi \;\; \mathrm{Tr}\left[\mathcal{E}(|\psi\rangle\langle\psi|)U(|\psi\rangle\langle\psi|)U^\dagger\right] \\ &= 1 - \int d\psi \;\; \mathrm{Tr}[\mathcal{E}(|\psi\rangle\langle\psi|)\mathcal{V}(|\psi\rangle\langle\psi|)] \end{aligned} \tag{2.45}$$

---

[2] The rigorous definition for unitarily-invariant probability measures involves many technical details which are not relevant for this thesis. Loosely speaking, the Haar measure is a probability measure η on the Borel subset of unitary operators Borel$(U(\mathcal{H}^n))$ that is unitarily invariant: $\eta(V\mathcal{A}) = \eta(\mathcal{A}) = \eta(\mathcal{A}V)$ for all $\mathcal{A} \in \mathrm{Borel}(U(\mathcal{H}^n))$ and $V \in U(\mathcal{H}^n)$. For the complete definition, see chapter 7.2 in [12].

which follows the cyclic property of trace. This measure is related to the fidelity function as

$$r(\mathcal{E},\mathcal{V}) = 1 - \int d\psi\ F(\mathcal{E}(|\psi\rangle\langle\psi|)\mathcal{V}(|\psi\rangle\langle\psi|))^2 \tag{2.46}$$

by recalling Equation 2.41 and that $\mathcal{V}(|\psi\rangle\langle\psi|)$ is a pure state if $\mathcal{V}$ is a unitary channel. Note that Equation 2.46 would not hold for a general CPTP map $\mathcal{V}$.

The average gate infidelity provides a way to quantify the "quality" of a quantum operation: if the gate we want to perform in the lab is represented by the ideal unitary channel $\mathcal{V}$ whereas the one actually carried out is represented by $\mathcal{E}$, then $r(\mathcal{V}^\dagger\mathcal{E}) = 1$ iff $\mathcal{V} = \mathcal{E}$ and $r(\mathcal{V}^\dagger\mathcal{E}) < 1$ otherwise [19].

In its original form, the average gate infidelity is not very easy to compute as it involves integration over the Haar measure. Fortunately, there exists a simple formula for r, using the following theorem[19, 20].

Theorem 2.5.3. *The average gate fidelity $\bar{F}$ for a TP map $\mathcal{E} \in T(\mathcal{H}^d)$ can be written as*

$$\bar{F}(\mathcal{E}) = \frac{dF_e(\mathcal{E}) + 1}{d + 1} \tag{2.47}$$

where

$$F_e(\mathcal{E}) \coloneqq \langle\phi|\,(\mathbb{1}_{L(\mathcal{H}^d)} \otimes \mathcal{E})(|\phi\rangle\langle\phi|)\,|\phi\rangle \tag{2.48}$$

is called the entanglement fidelity [21] for $\mathcal{E}$ with $|\phi\rangle$ being a maximally entangled state.

The second distance measure for quantum operations evaluates the worst-case distinguishability between two maps. First, we shall define a norm in the space of quantum maps

Definition 2.5.7. The diamond norm for a map $\Phi \in T(\mathcal{H}^n)$ is defined as

$$\|\Phi\|_\diamond \coloneqq \left\|(\Phi \otimes \mathbb{1}_{L(\mathcal{H}^n))}\right\|_{1\to 1} \tag{2.49}$$

where

$$\|\Phi\|_{1\to 1} \coloneqq \max_{\|A\|_1 \leq 1} \|\Phi(A)\|_1, \quad A \in L(\mathcal{H}^n) \tag{2.50}$$

is defined as the 1 → 1 norm for a map Φ.

The diamond norm is derived from the 1-norm over an extended space with twice the size as $L(\mathcal{H}^n)$ as its optimum value. The extended space gives rise to many useful properties of this norm: here we omit the proofs and list some of them as follows [12].

1. For all maps $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$, it holds that

$$\|\Phi\|_\diamond = \max_{u,v}\{\left\|(\Phi \otimes \mathbb{1}_{L(\mathcal{H}^n)})(uv^\dagger)\right\|_1 \ : \ u, v \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H})\} \tag{2.51}$$

   where $\mathcal{S}$ denotes the unit sphere, $\mathcal{S}(\mathcal{H}^n) = \{u \in \mathcal{H}^n : \|u\| = 1\}$.

2. For all channels $\Phi \in \mathcal{C}(\mathcal{H}^n, \mathcal{H}^m)$, it holds that $\|\Phi\|_\diamond = 1$.

3. For all maps $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$ and $U_0, V_0 \in U(\mathcal{H}^n),\ U_1, V_1 \in U(\mathcal{H}^m)$, it holds that

$$\|U_1\Phi(U_0XV_0)V_1\|_\diamond = \|\Phi\|_1 \tag{2.52}$$

4. *(Submultiplicativity)* For all maps $\Phi \in T(\mathcal{H}^n, \mathcal{H}^m)$ and For all maps $\Psi \in T(\mathcal{H}^m, \mathcal{H}^p)$, it holds that

$$\|\Psi\Phi\|_\diamond \leq \|\Psi\|_\diamond\|\Phi\|_\diamond \tag{2.53}$$

5. *(Multiplicativity to tensor product)* For all maps $\Phi_0 \in T(\mathcal{H}^n, \mathcal{H}^m),\ \Phi_1 \in T(\mathcal{H}^p, \mathcal{H}^q)$, it holds that

$$\|\Phi_0 \otimes \Phi_1\|_\diamond = \|\Phi_0\|_\diamond\|\Phi_1\|_\diamond \tag{2.54}$$

The metric induced by the diamond norm in the space of operators is called, as expected, the diamond norm distance.

Definition 2.5.8. The *diamond norm distance* between two maps $\Phi, \Psi \in T(\mathcal{H}^n, \mathcal{H}^m)$ is

$$d_\diamond(\Phi, \Psi) := \|\Phi - \Psi\|_\diamond \tag{2.55}$$

Similar to the average gate infidelity, one may define a single-argument version of the diamond norm distance between a noise channel $\mathcal{E} \in T(\mathcal{H}^n)$ and the identity channel $\mathbb{1}_{L(\mathcal{H}^n)}$, also known as the *worst case error rate*, as:

$$\epsilon(\mathcal{E}) := \frac{1}{2}d_\diamond(\mathcal{E}, \mathbb{1}_{L(\mathcal{H}^n)}) \tag{2.56}$$

Just like the role 1-norm played in the Holevo-Helstrom theorem, the diamond norm distance quantifies the optimal distinguishability between two quantum channels, as the following theorem states [12]:

Theorem 2.5.4. *Let* $\Phi_0, \Phi_1 \in \mathcal{C}(\mathcal{H}^n)$ *and* $\lambda \in [0,1]$*. For any choice of* $p \in \mathbb{Z}^+$*, a measurement* $\mathcal{M} = \{M_0, M_1\}$*, and a density operator* $\sigma \in \mathbb{D}(\mathcal{H}^n \otimes \mathcal{H}^p)$*, it holds that*

$$\lambda\langle M_0, (\Phi_0 \otimes \mathbb{1}_{L(\mathcal{H}^n)})(\sigma)\rangle + \langle(1-\lambda)M_1, (\Phi_1 \otimes \mathbb{1}_{L(\mathcal{H}^n)})(\sigma)\rangle \leq \frac{1}{2}(1 + \|\lambda\Phi_0 - (1-\lambda)\Phi_1\|_\diamond) \quad (2.57)$$

The above theorem implies that the optimal probability to distinguish two quantum channels is related to the diamond norm distance between them. We will re-examine the applicability of this theorem in Chapter 4 when the form of Φ_0 is obtained experimentally, whereas Φ_1 is some ideal target operation.

Chapter 3

How to Benchmark a Quantum Computer?

Quantum computing is deemed useful largely due to the rich complexity of quantum states and quantum processes. On the other hand, it is not hard to understand that characterizing quantum computers also becomes difficult due to these complexities. It is thus crucial to balance between the efficiency of the characterization protocol and the accuracy and amount of information obtained from the protocol. In this Chapter, we give a brief review of the commonly used protocols that aim to probe the performance of QIP devices from different approaches.

3.1 Quantum Tomography

As we have seen, quantum states and measurements can be described by positive operators, and quantum processes can be characterized by superoperators; thus the most straightforward method to learn about the performance of a quantum computer is to simply measure all these elements. This is the idea behind quantum tomography. The history of quantum state tomography can be dated back to the 19th century when Stokes derived the 4 famous equations that would fully determine the photonic polarization degree of freedom [22]. This idea of fully determining the state parameters was later generalized to other quantum systems, and was given the name quantum state tomography. There is an analogy between Stokes' design and more general state tomography protocols because coherent light beams can be viewed as an ensemble of two-level systems, with the two levels corresponding to the two polarization degrees of freedom. The Stokes parameters effectively allows one to reconstruct the density matrix of such a system[23].

Lying at the center of most tomography protocols is Born's rule, Equation 2.4. Assuming that multiple copies of the system exist, the experimentalist makes different measurements to the system and collect the frequencies for each outcome; these are then arranged in a data matrix, and information regarding the initial state may be gained through different methods, discussed in the next chapter. An important requirement on the measurements is that they must be informationally complete:

Definition 3.1.1. A set of measurements are said to be *informationally complete* if the outcome probabilities from this measurement completely determines the state of a quantum system.

It can be seen that in Equation 2.4, the role of states and measurement operators are symmetric: based on the same principle one can also probe an unknown measurement device using a set of well-characterized input states, a protocol which is termed detector (or measurement) tomography [24]. This technique is most widely used in quantum optics since the Q-function of a detector is directly measurable using coherent state inputs[1] [25, 26]. It is, in fact, enough to reconstruct all state preparation and measurement (SPAM) parameters if there is enough individual information on states and measurements. More formally, a set of SPAM $\{M_i, \rho_j\}$ is called *self-calibrating* if all possible detection probabilities

$$p_{ij} = \mathrm{Tr}\{M_i \rho_j\} \tag{3.1}$$

uniquely determines every element in $\{M_i, \rho_j\}$ [27]. Having a set of well-characterized self-calibrating SPAM allows one to reconstruct an unknown state directly from the measurement outcomes without the need to reconstruct the POVM elements first, provided that the statistics from the same measurements on these standard states are known. Both of the names *relative tomography* and *data pattern tomography* describes this group of protocols[28, 29]. Many realistic aspects, including how to choose wisely the set of self-calibrating states, have been studied in prior work (see also the experimental papers, e.g. [30]).

If an experimentalist has a set of well-characterized states as well as detectors, it is then possible to probe an unknown quantum process by inputting known states and making measurements, a protocol called quantum process tomography [31, 32]. It can be viewed

---

[1] Coherent states in optics are coherent superpositions of Fock states: $|\alpha\rangle = \sum_0^\infty \alpha^n \frac{e^{-|\alpha|^2/2}}{\sqrt{n!}} |n\rangle$

as a straightforward extension to state and measurement tomography: since a quantum operation can be described by a linear map, its effect can be completely determined provided that the action on a spanning set of input states is known. Process tomography has become a standard tool nowadays whenever full information about a quantum process is desired, and was also implemented for many common quantum gates as well as some algorithms such as quantum Fourier transform on many different implementations [33, 34, 35, 36, 37].

While these tomographic approaches give full information regarding the process, they are not scalable: for the combined system of n d-level sub-systems, the degrees of freedom (DoF) grows as d^n, which is exponential in the number of sub-systems. Thus, one requires exponentially many measurements to completely probe the state. The DoF for a general process happening on this combined system is on the order of d^{2n}, which is even larger and also scales exponentially. Thus, it is highly impractical to perform a full state tomography on a large system, not to mention a general process tomography. Another disadvantage shared by the aforementioned protocols is that they make assumptions about their "probes": for example in state tomography, it is assumed that the measurements are perfectly described by a set of completely known POVM elements. This is of course non-ideal, since all of states, gates and measurements are components of the same experimental setup that each requires characterization on their own.

3.2 Tomographic Reconstruction

Given the experimental frequencies, there are different ways to reconstruct the representations for states and processes. The most direct method is to invert Equation 2.4, and this is given the name linear inversion (LI) tomography. This can be done straight-forwardly using the Liouville representation: for state estimation, combining Equation 2.4 and 2.20 we can write the state as

$$|\rho\rangle\rangle = \sum_{j}^{4^n} \langle\langle P_j|\rho\rangle\rangle |j\rangle\rangle \tag{3.2}$$

Arranging the measured frequencies as a vector,

$$\vec{f} \coloneqq \sum_{i}^{r} \langle\langle M_i|\rho\rangle\rangle |i\rangle\rangle \tag{3.3}$$

where r is the number of independent measurements that were made. Inserting the identity written in terms of the Pauli basis vectors gives

$$\vec{f} \coloneqq \sum_{ij} \langle\langle M_i | P_j \rangle\rangle \langle\langle P_j | \rho \rangle\rangle | i \rangle\rangle \coloneqq A | \rho \rangle\rangle \tag{3.4}$$

where

$$A \coloneqq \sum_{ij} \langle\langle M_i | P_j \rangle\rangle E_{i,j} \tag{3.5}$$

When the number of informationally-complete measurement outcomes is equal to the dimension of the state space, $r = 4^n$ and the matrix A must be full-rank; one can therefore obtain the estimate

$$|\rho\rangle\rangle = A^{-1} \vec{f} \tag{3.6}$$

If the set of measurement outcomes can be over-complete, $r > 4^n$, the matrix $A^\dagger A$ will be invertible, so a least-squares estimate of the state can be calculated using the pseudo-inverse method as

$$|\rho\rangle\rangle_{LS} \coloneqq A^+ \vec{f}, \; A^+ \coloneqq (A^\dagger A)^{-1} A \tag{3.7}$$

Measurement and process tomography can be performed in a similar manner. For example, for process tomography, arranging the measured frequencies in a matrix and inserting two identity operators as before gives

$$\begin{aligned} \mathbf{f} &= \sum_{kl} \langle\langle M_k | G_\Phi | \rho_l \rangle\rangle | k \rangle\rangle \langle\langle l | \\ &= \sum_{ijkl} \langle\langle M_k | P_i \rangle\rangle \langle\langle P_i | G_\Phi | P_j \rangle\rangle \langle\langle P_j | \rho_l \rangle\rangle | k \rangle\rangle \langle\langle l | \\ &= A G_\Phi B \end{aligned} \tag{3.8}$$

where

$$A \coloneqq \sum_{ki} \langle\langle M_k | P_i \rangle\rangle | k \rangle\rangle \langle\langle i |, \; B \coloneqq \sum_{jl} \langle\langle P_j | \rho_l \rangle\rangle | j \rangle\rangle \langle\langle l | \tag{3.9}$$

For simplicity consider the case in which both A and B are square and full-rank, the estimate for G_Φ is given by $G_\Phi = A^{-1} \mathbf{f} B^{-1}$.

Linear inversion method has one most serious problem: it can easily give "unphysical" estimates under statistical fluctuations such as (unavoidable) counting errors, particularly so when the state under estimation is near the boundary of physically plausible states; in other words, close to pure states. Under such cases, the output from linear inversion tomography is likely to violate the usual physicality constraints such as being positive semidefinite. It is estimated that for low entropy and highly entangled states, this happens about 75% of the time [23]. Unfortunately, this is in many case the experimentally relevant situation, since many current quantum computing protocols require pure states as initial inputs. As a result, the output directly from a linear inversion tomography is unlikely to be very useful for referring future probabilities.

In 1996, the method of maximum likelihood estimation (MLE) applied to state tomography was first introduced by Hradil in [38]. This method ensures that the outputs satisfy the canonical physicality constraints (as defined in Chapter 2.3) by picking a member within the set of physical states that maximizes the "likelihood" that this state would produce the observed results. Specifically, given the measurement operators $\mathcal{M} = \{M_i\}$ and the corresponding observed occurrences n_i, and assuming that every outcome occurs independently from the others so that the total probability is the product of individual ones, one defines the likelihood function as

$$\mathcal{L}(\mathcal{M}|\rho) = \prod_i \mathrm{Tr}[M_i\rho]^{n_i} \tag{3.10}$$

which can be interpreted as the "likelihood that the measurement outcomes of $\mathcal{M}$ corresponds to the experiment, provided that the state being measured is ρ". Because of the n_i's on the exponent, if the total number of experiments is large, this function is often very sharply peaked at the maximum value and will be almost 0 anywhere else. This can create a problem when numerically optimizing over the set of ρ's. A solution to this problem is to bring down n_i from the exponent by transforming the above into a log-likelihood function, whose maximum occurs at the same position as the original function:

$$\log(\mathcal{L}(\rho|\mathcal{M})) = \sum_i n_i \log(\mathrm{Tr}[M_i\rho]) \tag{3.11}$$

Given the set of n_i, one can then maximize the log-likelihood within the set of valid density operators to obtain the one that "most likely" produces the data. Of course, MLE is not only restricted to state estimation: the same concept may be applied to measurement[39] and process[40, 33] tomography (see also more recent experiments [41, 42, 43]). As expected, MLE is also more computationally-demanding than LI due to the optimization step required. Luckily, the log-likelihood function has a unique local maximum because it is concave in ρ: this is because every term in 3.11 is the log of a non-negative and linear function in ρ, which is concave, and the sum of concave functions are concave[44]. The concavity of the function makes optimization easier and many standard algorithms exist that guarantees convergence to a unique answer (see, for example, [23, 45]). There are also diluted methods which promises to improve the convergence rate by adaptively changing the step-size parameter in every iteration[46].

MLE has it own problems: take MLE for state estimation as an example, it can be proved that when LI tomography outputs estimates that satisfies the canonical constraints where $\rho \in \mathbb{D}(\mathcal{H}^n)$ and $0 \leq M_i \leq \mathbb{1}_{L(\mathcal{H}^n)}$[2], the estimate from MLE would be identical to the LI estimate[39]. However, when LI produces a non-physical state (for example one with negative eigenvalues) which is outside of $\mathbb{D}(\mathcal{H}^n)$, then due to the convexity of the set of quantum states, the one that maximizes the log-likelihood function will always lie on the boundary of $\mathbb{D}(\mathcal{H}^n)$[44]. For single qubits this means that the estimate is a pure state; for multiple qubits, the estimate will contain zero eigenvalues which makes the reconstructed state rank-deficient. It was argued that zero eigenvalues are almost as bad as the negative ones produced by LI, since they essentially predict the occurrence of certain measurement outcomes to be impossible from only finitely many observed data points[44]. Since the goal of tomography is to generate models which can faithfully predict future measurement outcomes, this issue with MLE makes its predictions less credible.

Fundamentally, both LI and MLE correspond to a frequenist approach to interpreting probability, which takes the frequency from finite outcomes as the best approximate to the real probabilities. Recently, tomography with Bayesian methods were considered to be a better alternative over traditional frequenist methods by some authors[44, 47, 48]. In general, Bayesian estimation treats measured frequencies not as the best approximation to true probabilities, but as an update rule on a prior distribution $p(\rho)$ which represents our knowledge on the state before this measurement outcome (usually chosen to be as uninformative, or "flat", as possible in the absence of any measurement outcome). After every round of measurement, the posterior state is updated according to Bayes rule as $p(\rho|\mathcal{R}) \propto \mathcal{L}(\mathcal{M}|\rho;\mathcal{R})p(\rho)$ where $\mathcal{R}$ stands for a set of measured results. Instead of outputting a single state as MLE, Bayesian estimation outputs the most possible *probability distribution* based on currently available data. A point-estimate can then be generated from, e.g., the mean over this distribution, giving rise to the method of *Bayesian Mean Estimation* (BME). Bayesian methods naturally offers a complete description of one's current knowledge and is logically more in-line with the idea of state estimation. The advantage of BME over MLE is two-fold: first, it does not yield a rank-deficient estimate provided that the prior distribution is *robust* as defined in [44], in the sense that no finite data will completely rule out any possible outcomes. The robustness is a rather weak restriction: for example, the commonly used Haar measure over SU(d) satisfies this condition. Sec-

---

[2] The notation $A \geq 0$ implies $A \in \mathrm{Pos}(\mathcal{H}^n)$, and $A \geq B$ implies $(A-B) \in \mathrm{Pos}(\mathcal{H}^n)$, for $A,\ B \in L(\mathcal{H}^n)$.

ond, *error bars* on the estimate can be produced based on the standard deviation in the posterior distribution.

The most significant issue that has kept BME from being widely used in state estimation is its even higher demand of computational power over MLE. The essence of BME lies at calculating the posterior $p(\rho|\mathcal{R})$, which boils down to evaluating a (generally high-dimensional) integral of the likelihood function $\mathcal{L}$. This integral is usually analytically intractable and one would typically use approximating methods such as Markov-Chain Monte-Carlo[47]. More efficient sampling techniques have been explored in further detail in [47, 49, 50]

3.3 Solution: Randomized Benchmarking and Gate-Set Tomography?

As discussed earlier, the exponential scaling of quantum tomography makes full characterization of systems larger than a few qubits practically impossible, independent of the reconstruction methods. This scaling problem is even more significant for quantum processes. Moreover, unrealistic assumptions about ideal SPAM make estimates returned from conventional process tomography less credible. In order to overcome these difficulties, two different approaches may be taken: the first is to develop scalable methods that gains partial information about quantum gates; the second is to develop self-consistent methods that remove the assumptions about SPAM.

3.3.1 Randomized Benchmarking

The most widely-accepted protocol that belongs to the first category is randomized benchmarking (RB), along with its multiple variants[51, 52, 53, 54, 55, 56, 57, 58].

To put simply, RB is based on the fact that the overall quality of combining quantum gates decreases as the sequence length increases. The RB protocol is particularly suited for a set of gates $\mathcal{G}$ that forms a unitary 2-design (e.g., the Clifford group). We give the definition of the Clifford group and unitary t-designs below.

Definition 3.3.1. The Clifford group $\mathcal{C}_n$ on n-qubits is defined as

$$\mathcal{C}_n = \{V \in U(\mathcal{H}^{2^n}) \mid V\sigma V^\dagger \in \pm\mathbf{P}_n^* \ \forall\sigma \in \pm\mathbf{P}_n^*\}/U(1) \tag{3.12}$$

where $\mathbf{P}_n^* = \mathbf{P_n}\backslash\{I^{\otimes n}\}$ and $\mathbf{P}_n$ is the n-qubit Pauli group. In the definion, $/U(1)$" implies that equality is up to a global phase.

In other words, the Clifford group is the *normalizer* of $\mathbf{P}_n$ in $U(\mathcal{H}^{2^n})$.

Definition 3.3.2. A unitary t-design is a set of K unitary operators $\{U_j\}$ such that

$$\frac{1}{K}\sum_j p(U_j) = \int dU\ p(U)\ \forall p \in \mathrm{Hom}_{t,t}(U(d)) \tag{3.13}$$

where $\mathrm{Hom}_{t,t}(U(d))$ is the space of polynomials homogeneous of degree t in the matrix elements of both U and U^* (the complex conjugate of U). The integral is taken with respect to the normalized Haar measure in dimension d.

In fact, the Clifford group $\mathcal{C}_n$ is equivalent to the set of all possible ways of applying the Hadamard operator H, the phase operator P, and the 2-qubit CNOT operator (all being unitary) in an n-qubit system, and therefore is said to be generated by these three operations:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},\ P = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix},\ CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.14}$$

In the following text, we will use G to denote a target quantum channel, use $\tilde{G}$ to denote the real version of a channel, and write composition of channels as products of G's. The standard Clifford RB protocol can be summarized (in words) as follows[3][55, 56, 59, 57, 60, 61]:

1. Generate a random gate sequence $G_{m+1}G_m...G_1 \coloneqq G_{m+1:1} \in \mathcal{G}^{m+1}$ with length $m+1$ where each gate is chosen uniformly from the set of Clifford gates, and G_{m+1} equals to the inverse of the first m gates in the absence of gate errors;

2. Estimate the final survival probability $pr(G_{m+1:1}) = \mathrm{Tr}[E_0\ G_{m+1:1}(\rho)]$ by repeating the same sequence and record the outcome frequency corresponding to E_0;

---

[3] It is typically assumed that in the absence of SPAM errors, one of the projective measurement operators coincide with the initial state: $E_0 = \rho = |\psi\rangle\langle\psi|$

3. Repeat steps 1-2 for different random circuits of the same length m, evaluate the expectation value over $\mathcal{G}^{m+1}$, $\mathbb{E}_{\mathcal{G}^{m+1}} pr(G_{m+1:1})$, by randomly sample sequences with length $m+1$;
4. Repeat steps 1-3 for different values of m, fit to the decay model

$$\mathbb{E}_{\mathcal{G}^{m+1}} pr(G_{m+1:1}) = A + Bf^m \tag{3.15}$$

(where A, B are constants) to obtain the fit parameter which we denote as $\hat{f}$.

Standard Clifford group RB is scalable in the number of qubits n, as a result of the following theorem[62, 63]:

Theorem 3.3.1. *(Gottesman-Knill) Every (uniform family of) Clifford circuit(s), when applied to the input state $|0\rangle^{\otimes N}$ and when followed by a Z measurement of the first qubit, can be efficiently simulated classically.*

Thus, the crucial step in experimental design of standard RB, namely calculating the final gate to invert the m previously applied gates, can be efficiently performed on a classical computer. Earlier theoretical analysis (and more recent experimental realization) of RB protocols are based primarily on characterizing gates from the Clifford group due to this provable scalability in the number of qubits [55, 64, 5], although extensions to arbitrary elements in the unitary group has also been proposed [58]. Another important practical consideration is that although $|\mathcal{G}^m|$ also grows exponentially with the sequence length m, it is sufficient to randomly sample a relatively small subset of the whole space of all sequences in order to obtain a good estimate about the expectation value $\mathbb{E}_{\mathcal{G}^m} p_{\vec{G}}$, because the sample variance over $\mathcal{G}^m$ remains small for long sequences [57, 65].

Judging from how RB experiments are constructed, one expects that it would estimate the average quality of implementing a group of quantum gates. Indeed in an experiment, the estimated decay rate $\hat{p}$ can be used to calculate what's called the *RB number*:

$$r_{RB} \coloneqq \frac{(d-1)(1-\hat{f})}{d} \tag{3.16}$$

which serves as the final estimate of error for this set of gates. It has recently been shown that RB experiment with arbitrary gate-dependent but Markovian noise will always produce a single exponentially decaying curve, with an additional term accounting for the gate-dependence that decays exponentially with the sequence length [60, 61]. This agrees with the results from multiple experiments where a single exponential decay curve was observed [5, 66, 67, 64, 68], despite the fact that the actual errors occurring in an experiment are almost always gate-dependent. This justifies the practice of extracting a single RB number from most experiments.

The RB number is undoubtedly real and operational: the larger it is, the worse the device is performing overall. On the other hand, there have been difficulties in interpreting this number and relating it to conventional figures of merit. It has been proved that if the error is gate-independent, i.e., $\tilde{G}_i = \mathcal{E} G_i \; \forall i$ where $\mathcal{E}$ is an error channel, then $r_{RB} = r(\mathcal{E})$ (as defined in Equation 2.43) is exactly the average infidelity of the error channel $\mathcal{E}$[55]. And for gate-dependent errors, it has long been argued in many RB literature that $r_{RB} \approx \mathbb{E}_i r(\mathcal{E}_i)$, i.e., the average of the average infidelity over all error channels. However, it was recently shown in numerical simulations that these two numbers can differ by orders of magnitude [61], and thus $\mathbb{E}_i r(\mathcal{E}_i)$ is not a valid interpretation for r_{RB}. The reason is related to the freedom in representing quantum operations, which is the central topic of the next chapter.

3.3.2 Gate-Set Tomography

While RB provides a protocol which is scalable in the number of qubits, it does not provide tomographic information on the operations. Gate-Set Tomography (GST) [16, 69, 70, 71] was an important attempt corresponding to the second approach above, which aims at performing self-consistent SPAM *and* gate tomography without excessive assumptions like those mentioned at the beginning of this section. The assumptions made in a GST experiment include[69]:

1. The Hilbert-space dimension n is known;
2. Initialization can repeatedly prepare the system in the same state ρ;
3. The measurement can be described by the POVM $\{M_1, M_2 = \mathbb{1}_{\mathcal{H}^n} - M_1\}$;
4. Every gate can be described by a CPTP map in $T(\mathcal{H}^n)$.

which are much weaker compared to those in conventional process tomography.

A gate-set is a mathematical description of all the possible actions executable in an experiment, typically consisting of models for initial states ($\mathcal{S}$), gate operations ($\mathcal{G}$), and measurements ($\mathcal{M}$), written together as $\mathbb{G} = \{\mathcal{S}, \mathcal{G}, \mathcal{M}\}$. One may choose to write elements in each of these subsets using different representations such as the ones introduced in Chapter 2. GST takes into consideration that the number of initial preparable states and possible measurements on a current quantum computer is limited (usually just one state $|0\rangle\langle 0|$ and a 2-outcome projective measurement), so the spanning set of states/measurements are created by performing gates to transform them into different ones. And since these gates are precisely the ones being characterized in a process tomography, it is possible to self-consistently characterize them.

The reconstruction steps of GST are largely similar to those of a conventional state/process tomography experiment. We briefly demonstrate the idea with Linear GST [69, 16]: for simplicity, we consider a setup where the initial state consists of one element and the measurement is a two-outcome POVM. First, one can transform the initial state $|\rho\rangle\rangle$ using an arbitrary combination of gates in $\mathcal{G}$: for Linear GST to work, this set must generate an complete set of states and measurements. We denote this set as the SPAM gate strings $\mathcal{F} = \{F_1, ..., F_N\}$, with each member being composed of gates in $\mathcal{G}$,

$$F_k = G_{f_{k_1}} \circ G_{f_{k_2}} \circ ... \circ G_{f_{k_{L_k}}} \tag{3.17}$$

where $\{f_{k_l}\}$ are indices labeling gates in $\mathcal{G}$ and L_k is the length of the $k-$th SPAM string. All possible probabilities given these SPAM gates can be expressed as

$$p_{ijk} = \langle\langle M_1|F_i G_k F_j|\rho\rangle\rangle \tag{3.18}$$

Like in Equation 3.8 we insert two identity operators,

$$p_{ijk} = \sum_{rs} \langle\langle M_1|F_i|r\rangle\rangle\langle\langle r|G_k|s\rangle\rangle\langle\langle s|F_j|\rho\rangle\rangle := \sum_{rs} A_{ir}(G_k)_{rs}B_{sj} \tag{3.19}$$

where A and B contains information about the SPAM as

$$\begin{aligned} A &= \sum_i |i\rangle\rangle\langle\langle M_1|F_i \\ B &= \sum_j F_j|\rho\rangle\rangle\langle\langle j| \end{aligned} \tag{3.20}$$

which are unknown by the GST assumption. However, the experimentally measured values p_{ijk} corresponds to the (i, j)-th component of the matrix

$$\tilde{G_k} := A G_k B \tag{3.21}$$

and WLOG one may assume $G_0 = \mathbb{1}_{L(\mathcal{H}^n)}$ which gives

$$\tilde{G}_0 = AB \tag{3.22}$$

Thus

$$\hat{G}_k \coloneqq \tilde{G}_0^{-1}\tilde{G}_k = B^{-1}A^{-1}AG_kB = B^{-1}G_kB \tag{3.23}$$

is an estimate for the gate G_k, up to a similarity transform by an unobservable B. The states and measurement vectors can be estimated by performing another set of experiments with $G_k = \mathbb{1}$, and collecting the outcomes as (component-wise identical) vectors

$$\begin{aligned} |\tilde{\rho}\rangle\rangle &\coloneqq \sum_i |i\rangle\rangle\langle\langle M_1|F_i|\rho\rangle\rangle = A|\rho\rangle\rangle \\ \langle\langle\tilde{M}_1| &\coloneqq \sum_i \langle\langle M_1|F_j|\rho\rangle\rangle\langle\langle j| = \langle\langle M_1|B \end{aligned} \tag{3.24}$$

so estimates for SPAM can be computed as

$$\begin{aligned} |\hat{\rho}\rangle\rangle &\coloneqq \tilde{G}_0^{-1}|\tilde{\rho}\rangle\rangle = B^{-1}|\rho\rangle\rangle \\ \langle\langle\hat{M}_1| &\coloneqq \langle\langle\tilde{M}_1| = \langle\langle M_1|B. \end{aligned} \tag{3.25}$$

This estimate is not equivalent to the original gate-set; instead it is transformed by

$$|\tilde{\rho}\rangle\rangle = B^{-1}|\rho\rangle\rangle,\ \langle\langle\tilde{M}_1| = \langle\langle M_1|B,\ \tilde{G}_k = B^{-1}G_kB \tag{3.26}$$

But the claim is that this estimate is as good as the original gate-set in describing the dynamics of this set of operations, as it gives identical prediction about the outcome probabilities from all possible experiments using only these operations, since

$$\langle\langle\tilde{M}_1|\tilde{G}_k|\tilde{\rho}\rangle\rangle = \langle\langle M_1|G_k|\rho\rangle\rangle \tag{3.27}$$

This freedom in representations using different gate-sets is called *gauge equivalence* in the literature [69, 72, 73].

Although GST failed to produce an exact representation of the gate-set elements, in practice some methods that aims to recover the original representations are applied. For example, one may gauge-transform the estimated gate-set by some invertible matrix $\tilde{B}$ as

$$|\rho'\rangle\rangle = \tilde{B}^{-1}|\tilde{\rho}\rangle\rangle,\ \langle\langle M_1'| = \langle\langle\tilde{M}_1|\tilde{B},\ G_k' = \tilde{B}^{-1}G_k\tilde{B} \tag{3.28}$$

and try to make this transformed gate-set look as close to the target as possible by some pre-defined distance measure between gate-sets, a practice known as gauge optimization [71]. Consequence of this act will be discussed at the end of next chapter. GST can be useful in other aspects than producing tomographic estimates: for example, it was claimed to be useful for detecting non-Markovianity in the gates by implementing and measuring long gate strings [71].

Chapter 4

Gauge in Circuit Model

In the previous chapter, we saw how methods like standard state or process tomography make assumptions about the other components being used to perform the experiment. We also illustrated with GST experiments how a completely self-consistent protocol fails to give an exact description of a gate-set, and argued that this is due to a gauge degree of freedom (DoF) in quantum theory. Despite its conceptual importance in representing quantum operations, this DoF has mostly been overlooked when tomography experiments are reported. Only few theoretical analysis were conducted that pays particular attention to this subject [61, 73, 72]. In this chapter, we clarify the concept regarding this gauge, and demonstrate some of its implications on interpreting the gate-set representation for quantum objects. In Section 4.1 we analyze an experimentally-motivated example which demonstrates the effects of this gauge DoF in a tomography experiment. In Section 4.2 we give definitions for gauges and gauge transformations, as well as their role in representing quantum states as mathematical objects. In Section 4.3 we discuss the implications brought by this DoF, in particular addressing why distance metrics like the diamond norm distance between a measured gate and a target loses its operational meaning. We also discuss several practices that are commonly used and yet are unhelpful regarding this problem. In Section 4.4, we suggest a different measure of quality for gate-sets which is gauge-invariant and well-defined. We then provide an experimentally viable protocol that can be useful in calibrating quantum devices in a way that is solely based on measurable quantities, thereby fully operational.

4.1 Example

To see the emergence of this gauge naturally, we will again represent components of a QIP device in the gate-set picture: recall that a gate-set is a mathematical description of all possible actions executable in an experiment, typically consisting of models for initial states ($\mathcal{S}$), gate operations ($\mathcal{G}$), and measurements ($\mathcal{M}$). For example, if in a particular experiment (in the absence of errors) a person initialized a qubit in the state $|0\rangle$, applied a Z-rotation gate with angle θ, and performed a projective measurement in the $\{|0\rangle, |1\rangle\}$ basis, then the experimentalist's control can be represented by the following gate-set:

$$\mathbb{G} = \left\{ \mathcal{S} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\},\ \mathcal{G} = \left\{ \begin{pmatrix} 1 & 0 & 0 & e^{-2i\theta} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ e^{2i\theta} & 0 & 0 & 1 \end{pmatrix} \right\},\ \mathcal{M} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\} \right\} \tag{4.1}$$

where we have chosen to represent states as density operators ρ, a gate Φ in its Choi representation J_Φ, and measurements as measurement operators E. Now consider the following two gate-sets written in the same representation:

$$\mathbb{G}_1 : \begin{cases} \mathcal{S} = \left\{ \frac{1}{2} \begin{pmatrix} 1+\epsilon_1 & 0 \\ 0 & 1-\epsilon_1 \end{pmatrix} \right\}, \\ \mathcal{G} = \left\{ \frac{1}{2} \begin{pmatrix} 1 & iqe^{-\gamma} & iqe^{-\gamma} & q \\ -iqe^{-\gamma} & 1 & q & -iqe^{-\gamma} \\ -iqe^{-\gamma} & q & 1 & -iqe^{-\gamma} \\ q & iqe^{-\gamma} & iqe^{-\gamma} & 1 \end{pmatrix} \right\}, \\ \mathcal{M} = \left\{ \frac{1}{2} \begin{pmatrix} 1 & i\epsilon_2 \\ -i\epsilon_2 & 1 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 1 & -i\epsilon_2 \\ i\epsilon_2 & 1 \end{pmatrix} \right\} \end{cases} \quad \mathbb{G}_2 : \begin{cases} \mathcal{S} = \left\{ \frac{1}{2} \begin{pmatrix} 1+p\epsilon_1 & 0 \\ 0 & 1-p\epsilon_1 \end{pmatrix} \right\}, \\ \mathcal{G} = \left\{ \frac{1}{2} \begin{pmatrix} 1 & ipqe^{-\gamma} & i\frac{q}{p}e^{-\gamma} & q \\ -ipqe^{-\gamma} & 1 & q & -i\frac{q}{p}e^{-\gamma} \\ -i\frac{q}{p}e^{-\gamma} & q & 1 & -ipqe^{-\gamma} \\ q & i\frac{q}{p}e^{-\gamma} & ipqe^{-\gamma} & 1 \end{pmatrix} \right\}, \\ \mathcal{M} = \left\{ \frac{1}{2} \begin{pmatrix} 1 & i\epsilon_2 \\ -i\epsilon_2 & 1 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 1 & -i\epsilon_2 \\ i\epsilon_2 & 1 \end{pmatrix} \right\} \end{cases} \tag{4.2}$$

with the parameters

$$\epsilon_1 = \epsilon_2 = 0.9,\ \gamma = 0.1,\ p = \frac{10}{9},\ q = 0.8 \tag{4.3}$$

It can be verified that both states are positive-semidefinite and have trace 1, both sets of measurement operators are positive-semidefinite and sum up to identity, and both gates are completely-positive and trace-preserving (CPTP) because $J_\Phi \geq 0$ and $\mathrm{Tr}_2[J_\Phi] = \mathbb{1}_{\mathcal{H}^2}$, according to Theorem 2.3.1. Therefore, both are "valid gate-sets" satisfying the canonical constraints one would expect. The two gate sets appear to describe different experimental setups since their elements do not match: for example, the initial state in $\mathbb{G}_1$ is a mixed Z-state with polarization 0.9, whereas the one in $\mathbb{G}_2$ is a pure state. Assume now that one is given two black-box quantum computers, accessible only externally through "buttons", that are each described by one of the above two gate-sets (initialize a state in $\mathcal{S}$, execute a gate in $\mathcal{G}$, measure $\mathcal{M}$ and display the outcome)[69]. An experimentalist takes these two black-boxes and tries to distinguish between one another by recording the outcome statistics with arbitrary available gate combinations. It would then be discovered that the same protocol performed on both setups would give an identical outcome distribution, despite the fact that they are indeed described by two different gate-sets. Are these two quantum computers, then, performing in identical ways?

In order to naturally see why this happens, let's re-express the gate-set elements using the Liouville representation where states and measurement operators are represented as vectors, and quantum gates as super-operators (matrices), as introduced in Section 2.3.3. This representation is a linear bijection with the previous one and is thus unique (see Appendix section A for conversion formulas between Pauli and Choi representations). The gate-set elements can now represented as

$$\mathbb{G}_1 : \begin{cases} \mathcal{S} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & \epsilon_1 \end{pmatrix}^T \right\}, \\ \mathcal{G} = \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q & 0 & 0 \\ 0 & 0 & 0 & -qe^{-\gamma} \\ 0 & 0 & qe^{-\gamma} & 0 \end{pmatrix} \right\}, \\ \mathcal{M} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -\epsilon_2 & 0 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & \epsilon_2 & 0 \end{pmatrix} \right\} \end{cases} \qquad \mathbb{G}_2 : \begin{cases} \mathcal{S} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & p\epsilon_1 \end{pmatrix}^T \right\}, \\ \mathcal{G} = \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q & 0 & 0 \\ 0 & 0 & 0 & -\frac{q}{p}e^{-\gamma} \\ 0 & 0 & pqe^{-\gamma} & 0 \end{pmatrix} \right\}, \\ \mathcal{M} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -\epsilon_2 & 0 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & \epsilon_2 & 0 \end{pmatrix} \right\} \end{cases} \tag{4.4}$$

Recall that in the Liouville representation, Born's rule for the probability of a measurement outcome M_1 given an input state ρ becomes

$$\text{prob} = \langle\langle M_1 | \rho \rangle\rangle \tag{4.5}$$

and according to Equation 2.22, if t gates G_1,...,G_t are applied to the state before the measurement takes place, it is expressed as matrix multiplication resulting in an outcome probability of

$$\text{prob} = \langle\langle M_1 | G_t ... G_2 G_1 | \rho \rangle\rangle \tag{4.6}$$

Now it is not difficult to see that if the following transformation is applied to all the elements in a gate-set

$$|\rho\rangle\rangle \to B|\rho\rangle\rangle,\ \langle\langle M_1| \to \langle\langle M_1|B^{-1},\ G_\Phi \to BG_\Phi B^{-1} \tag{4.7}$$

for some invertible matrix B, then the outcome probabilities will remain unchanged for any combination of states, gates and measurements. Since these probabilities are the only experimentally-accessible quantities, this new gate-set describes an experimental setup that is indistinguishable from the previous one: the same experimental result can be interpreted equally well by these two gate-sets. This is the gauge DoF inherently present in mathematically representing quantum experiments, in analogy with concepts in thermodynamics and electromagnetism [74], with B being called the gauge transformation matrix (to be defined later). The analogy arises from the fact that changing the gauge does not result in observable effects in an experiment, just as changing the electromagnetic gauge would not result in any difference in the measurable electric or magnetic fields.

The reader may now verify that the two gate-sets in Equation 4.4 are indeed related by the gauge transformation matrix

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{4.8}$$

Moreover, the reader may verify that the gate-set $\mathbb{G}_1$ corresponds to an apparatus capable of preparing a mixed state $\rho = \frac{1}{2}(I + \epsilon_1 Z)$, making a POVM measurement $\{M_1, \mathbb{1}_{\mathcal{H}^2} - M_1\}$with $M_1 = \frac{1}{2}(I - \epsilon_2 Y)$, and performing a gate $\Phi(\rho) = \int d\theta e^{-i\theta X}\ \Phi_D(\rho, q)\ e^{i\theta X} \eta(\theta, \gamma)$ where $\eta(\theta, \gamma) = \frac{1}{\sqrt{\pi\gamma}} e^{-\frac{(\theta - \pi/4)^2}{\gamma}}$ is a Gaussian distribution and $\Phi_D(\rho, q) = (\frac{3q+1}{4})\rho + \frac{3(1-q)}{8} I$ is the depolarizing channel. The applied gate is a $\frac{\pi}{2}$ rotation along the x-axis with some normally distributed error on the actual rotation angle with the qubit being succumbed to a depolarizing channel. The gate-set $\mathbb{G}_2$ is a gauge transformed version of $\mathbb{G}_1$ according to Equation 4.7. Finally, to address our question about whether these two devices are performing identically, our answer would be that if they are performing differently but their behavior are related through a gauge transformation, it is not an experimentally-detectable difference.

While we have shown that equivalent gate-sets of the same experiment exist, is it even a cause for concern? The main issue with this gauge DoF arises when we try to evaluate meaningful metrics quantifying how good our control over our quantum system is. Most of the metrics for quantum states and gates depend on the specific form of the representation and are thus not gauge-invariant. For example, in the two representations $\mathbb{G}_1$ and $\mathbb{G}_2$, the initial states clearly have different Z polarizations and would be considered to be different states by any measure. The same applies for gate measures: for example, consider the diamond norm distance defined in Equation 2.55 between two quantum channels. We used the DiamondNorm function in the package QETLAB [75, 76] to compute this quantity for the gates in $\mathbb{G}_1$ and $\mathbb{G}_2$, and obtained $\|\Phi_1 - \Phi_0\|_\diamond = 0.376$ and $\|\Phi_2 - \Phi_0\|_\diamond = 0.372$, where Φ_0 is the gate in the "ideal" gate-set by taking $p = q = 1$ and $\gamma = 0$, confirming a difference in diamond norm for these two representations. While this difference appears to be small, it is sufficient to show that this problem exists in principle; in general, one can easily construct an example with a larger difference by applying a unitary, basis-rotating gauge transformation on a gate-set. The purpose of our example is to demonstrate that the freedom in a gauge-transformation can be more than a change of basis: for example, it can also change the purity of the initial state representation, which may cause deeper confusions in device characterization compared to a basis mismatch.

4.2 Gauge and Representation of quantum states

We have seen that under realistic circumstances, the same experiment can be described by different-looking gate-sets due to a gauge DoF. In this section we illustrate pictorially how representations of quantum states are related to the concepts of gauges and gauge transformations. For clarity we focus on the representation for quantum states, but similar arguments can be made about gates and measurement operations.

From the point of view of scientific realism, the apparatus (e.g., a qubit) has a physical existence and properties (which may be relative to the environment) independent of our representation. We describe the *abstract state of this physical object* as a *noumenal state* following the terminology in [77], denoted as N in Figure 4.1. Here, we slightly change their definition to include in N both physically allowed (denoted as P) and forbidden states. Quantum mechanics allows us to assign to each noumenal state a mathematical *representation* which is an element of a Hilbert space $\mathcal{H}^d$: for example, one can associate the system with a matrix that summarizes its properties, and the set of all $d \times d$ matrices is what we call R in the same figure. Such an association is what we call a *gauge* Γ, which is a bijective map from N to R: the bijectivity of the map should be clear from our inclusion of physically-forbidden states in N, which allows assigning "some state" to every $d \times d$ matrix. Different choices of Γ thus correspond to different descriptions of the noumenal states.

The common formulation of quantum mechanics says that every state of a quantum object can be described by a density operator [78], which belongs to a subset of the *canonical constraints* as defined at the beginning of Chapter 2.3. This means that there exists a *canonical gauge*

$$\Gamma_1 : \ N \to R, \ \Gamma_1(P) = I_1 \tag{4.9}$$

where $I_1 = \mathbb{D}^d$ is the set of $d \times d$ density operators. In fact, there exists a family of canonical

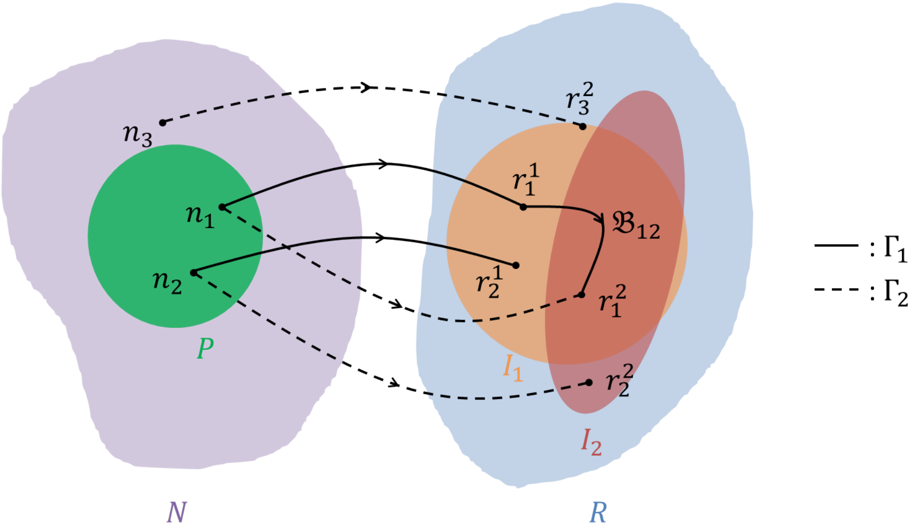

Figure 4.1: Diagram illustrating (in the operator picture) noumenal states, representations and gauges. N denotes the set of all noumenal states and P represents the set of states that are physically allowed to prepare in principle. Elements in N are denoted as n_i. R denotes the set of representations for the noumenal states, which can be the set of all 2×2 matrices for qubits. I_i is the image of P under gauge Γ_i, and r_i^j is the image of n_i under gauge Γ_j: in our example $I_1 = \mathbb{D}^d$ is the set of density operators in $L(\mathcal{H}^d)$. $\mathfrak{B}_{ij}$ represents a gauge transformation from Γ_i to Γ_j. Whether an object in R directly corresponds to objects in P or not depends on the particular gauge under which the object is represented.

gauges that are all related to Γ_1 through unitary gauge transformations, with some B that describe unitary transformations in $\mathcal{U}(\mathcal{H}^d)$: although each canonical gauge maps the same point in P to different points in I_1, the image of P remains unchanged. Satisfying the canonical constraint implies that we should work in one of these canonical gauges. Now, consider another gauge Γ_2 which can be converted from Γ_1 with a *gauge transformation* $\mathfrak{B}_{12}$, defined by

$$\mathfrak{B}_{12} \coloneqq \Gamma_2(\Gamma_1^{-1}),\ \mathfrak{B}_{12}(r_*^1) = r_*^2 \tag{4.10}$$

and in the light of Equation 4.7, can be represented in the Pauli basis as

$$|\mathfrak{B}_{12}(\rho)\rangle\rangle \coloneqq B_{12}|\rho\rangle\rangle,\ \langle\langle\mathfrak{B}_{12}(M)| \coloneqq \langle\langle M|B_{12}^{-1},\ G_{\mathfrak{B}_{12}(\Phi)} \coloneqq B_{12}G_\Phi B_{12}^{-1} \tag{4.11}$$

As a subset of R, I_1 is generally not invariant under an arbitrary gauge transformation. Consider a general trace-preserving transformation given by the following transformation matrix

$$B_{12} = \begin{pmatrix} 1 & 0 \\ \vec{x} & y \end{pmatrix} \tag{4.12}$$

where $\vec{x}$ is a $(d-1)$ by 1 real vector and y is a $(d-1)$ by $(d-1)$ real matrix: this affine transformation of I_1 will result in a different subset of R which we call I_2, the image of P under Γ_2. Such a gauge is perfectly valid in principle, provided that *all* the gates and measurement operators are transformed according to 4.11 as well, even though I_2 is no longer the set of density operators.

The existence of a non-canonical gauge implies, for example, that a physical state may or may not be represented by a density operator: as illustrated in Figure 4.1, $r_1^2 \in I_1$ whereas $r_2^2 \notin I_1$. Similarly, a density operator in a non-canonical gauge does not necessarily correspond to a physical state, as $r_3^2 \in I_1$ but $n_3 \notin P$. Now, using $B_{12} = B$ from Equation 4.8, the image of n_3 under Γ_2 becomes $\frac{1}{2}(I + \sigma_z)$, which *is* in I_1 but only as a consequence of the non-canonical gauge. We conclude that if the gauge is unknown, the mathematical form of the representations does not automatically imply the noumenal state being physically allowed or not. Of course, representations in a canonical gauge are much easier to work with, so traditionally in a tomography experiment one aims to express all gate-set elements in a natural gauge. But since the gauge is not directly available, and one can only change from one gauge to the other by means of gauge transformations, it is impossible to conclude the gauge being canonical or not when requiring only a finite number of tested states to satisfy the canonical constraints, without making assumptions about the particular form of any gate-set element.

4.3 Discussion

The existence of this gauge DoF has direct implications on quality measures for quantum operations. The main problem is that there is no way to know whether an experimentally-determined gate-set element is expressed in the same gauge as their target, which by definition are written in a canonical gauge. Since changing the gauge will change the expression for the experimental elements, any distance measure between the two will likely change upon changing this indeterminable gauge. We have already seen in Section 4.1 that by changing the gauge, the states can appear as having different expressions; the same holds true for gates and measurement operators.

From quantum information theory, we have successfully attached some operational meanings to various distance metrics: an important example is the interpretation for diamond norm distance (Equation 2.57) which can be simplified to

$$\|\Phi_0 - \Phi_1\|_\diamond = \frac{1}{2} \max_{0 \leq M_1 \leq \mathbb{1}_{\mathcal{H}^{d^2}}, \rho \in \mathbb{D}^{d^2}} \langle\langle M_1 | ((G_0 - G_1) \otimes \mathbb{1}_{L(\mathcal{H}^d)}) | \rho \rangle\rangle \tag{4.13}$$

where Φ_0 and Φ_1 are CPTP maps, and $\{M_1, \mathbb{1}_{\mathcal{H}^{d^2}} - M_1\}$ is a two-outcome measurement[12]. One can interpret Equation 4.13 as that $\|\Phi_1 - \Phi_0\|_\diamond$ corresponds to an optimal one-shot distinguishability between Φ_1 and Φ_0 by allowing arbitrary physical input states and making arbitrary measurements. However, in the situation where Φ_0 is an experimentally-measured gate and Φ_1 is its ideal target, the above interpretation that leads to its operational meaning becomes fundamentally flawed. The reasoning is as follows: in order to compute this norm one would need the representations for Φ_0 and Φ_1, for example their Choi representation. While Φ_1 is by definition represented in a canonical gauge, for Φ_0 the gauge is unknown. So the value of this norm could change if one artificially chooses to express Φ_0 in a different gauge, since its value depends on the exact representation. Next, consider the case where Φ_1 represents the identity gate, which has the same expression in any gauge. Under this case one may say with confidence that Φ_0 and Φ_1 can be considered to be in the same gauge; however, one must also be aware that (one out of the many equivalent ways of) computing this norm requires optimizing over a given set of states and measurements, whose representation is not gauge-invariant. By default, they are expressed in a canonical gauge. Since we do not know in an experiment the gauge in which the gates are represented, one typically assumes (implicitly) that Φ_0 is in a canonical gauge, use it as an input in 4.13, and perform the optimization according to the canonical constraints. Obviously, while one *wants* to obtain the value when everything is expressed in the same canonical gauge, they will typically not get this value if the optimization region is defined with respect to the canonical gauge while the function being optimized is not. Note that while a canonical gauge is in principle as valid as any other gauges, the desired value can only be obtained when all gate-set elements are expressed in the *same* gauge: this can be shown using the fact that for a change of variable $y = g(x)$,

$$\max_{x \in I} F(x) = \max_{y \in g(I)} F(g^{-1}(y)) \tag{4.14}$$

for all functions $F(x)$, so

$$\begin{aligned}\max_{\rho\in I_1, M_1\in J_1} \langle\langle M_1|G_{\Gamma_1}|\rho\rangle\rangle &= \max_{\sigma\in\mathfrak{B}_{12}(I_1), N_1\in\mathfrak{B}_{12}(J_1)} \mathfrak{B}_{12}^{-1}(\langle\langle N_1|)G_{\Gamma_1}\mathfrak{B}_{12}^{-1}(|\sigma\rangle\rangle) \\ &= \max_{\sigma\in I_2, N_1\in J_2} \langle\langle N_1|B_{12}G_{\Gamma_1}B_{12}^{-1}|\sigma\rangle\rangle \\ &= \max_{\sigma\in I_2, N_1\in J_2} \langle\langle N_1|G_{\Gamma_2}|\sigma\rangle\rangle\end{aligned} \tag{4.15}$$

where we used J_1 to denote the set of all operators σ satisfying $0 \leq \sigma \leq \mathbb{1}$. If only the gauge for gates is changed, the outcome will potentially change as well. And since its value is not unique with respect to experimental data, the "diamond norm distance" in this case cannot have a concrete operational meaning. Put in a different way, the outcome from such optimization protocols can only be attached a meaning when the elements are represented in the same gauge, an option which only exists in theoretical analysis.

In Chapter 3.3.1 and 3.3.2, we mentioned that the decay parameter r_{RB} in an RB experiment cannot be directly interpreted as an "averaged" average gate infidelity over different error channels, and that GST experiments cannot fully determine every gate-set element without prior assumptions on their forms. Both of these issues are in fact due to this gauge freedom in expressing quantum operations. The "problem" with GST shall be clear by reviewing Chapter 3.3.2 on how Linear GST is performed: since no form of SPAM is being assumed, only the matrix $\tilde{G}_0$ in Equation 3.22 is known instead of the individual matrices A and B, allowing a gauge transformation between state and measurement operators. For the interpretation of r_{RB}, the subtlety arises because $\mathbb{E}_i(r(\mathcal{E}_i))$ depends on the representation of errors on each gate, $\mathcal{E}_i$. While r_{RB} is real and measurable, $\mathcal{E}_i$ can change upon a gauge-transformation of the gate-set, which also changes the value of $\mathbb{E}_i(r(\mathcal{E}_i))$: therefore, $\mathbb{E}_i(r(\mathcal{E}_i))$ (the average of $r(\mathcal{E}_i)$ over all gates G_i) is not an operational quantity like r_{RB} and cannot be equal to each other.

We briefly discuss several common practices related to this DoF in quantum tomography. As we mentioned at the end of section 3.3.2, the process of "gauge optimization" is commonly adopted in GST experiments whereby the gauge transformation matrix B is varied to minimize the distance from the target gate-set according to a (non-gauge-invariant) weighted distance measure[71]. However, this optimized gauge is just as arbitrary as any other gauge, and the resultant gate-set is still not a faithful representation of the apparatus. Moreover, such optimization undermines a common use of tomography, namely, assessing the performance of a system against some external threshold (e.g., a fault-tolerance threshold). Altering the gauge to make the channel being tested look like the target will artificially reduce the distance between the two, and assigning different weights on SPAM and gates will result in a difference in the output even though such weights are only based on rough initial guess about the relative quality of these components. Another common approach is Maximum Likelihood Estimation (see section 3.2), which takes the estimated gate-set to be the one that maximizes the likelihood function of obtaining the experimental data, while restricting the gate-set elements to satisfy the canonical physicality constraints[79, 80]. However, it does not resolve the gauge ambiguity either because all gauge-equivalent gate-sets are equally likely to produce the data by definition. In the process of optimization, one will find that the likelihood function profile has the same value wherever two points are related by a gauge-transformation, and the actual output is largely a matter of the optimization algorithm and the initial parameters [81].

4.4 A gauge-invariant measure for gate-sets

We have seen that the gauge DoF prevents one from using conventional distance measures to faithfully evaluate the "goodness" of quantum operations. Note that our discussion is carried out in the absence of any additional errors such as finite-counting, and in a real experiment the situation becomes even more complicated. Fundamentally, this problem is due to the limited information that can be gained from experimental probabilities, compared to the full characterization in the gate-set picture. A gauge-transformation re-assigns state, gate, and measurement "errors" by adjusting their relative appearance in different representations, while keeping the experimental measurables unchanged. This implies that assumptions are somehow inevitable when representing real-life quantum systems in the gate-set picture. Of course, it should also be clear from our discussion that this DoF can be eliminated by asserting certain assumptions, for example by manually demanding the initialized state to be $|0\rangle$, or by reporting everything with respect to the measurement apparatus (i.e., assuming perfect measurements). Depending on the specific setup, the validity of these assumptions may largely vary.

As far as we know, no fully gauge-invariant measure between quantum gates has been reported [69, 81], nor for those between quantum states/measurements. In fact, what the previous sections have established is that such measures cannot exist, because representations for individual states/measurements/gates are intrinsically non-gauge-invariant. Below we propose a new, alternative distance measure for a *gate-set* which is gauge-invariant, and provide a protocol to characterize and improve experimental control based on this measure. As in section 4.1, let $\mathbb{G}$ denote the gate-set $\{\mathcal{S}, \mathcal{G}, \mathcal{M}\}$; let $\vec{C}_m$ denote a particular experiment with input state $\rho \in \mathcal{S}$, measurement $M \in \mathcal{M}$, and a set of m gates $G_1 ... G_m$ each selected from $\mathcal{G}$. A well-motivated quality measure for $\vec{C}_m$ is the total variation distance between the measured outcome probability distribution and the ideal one,

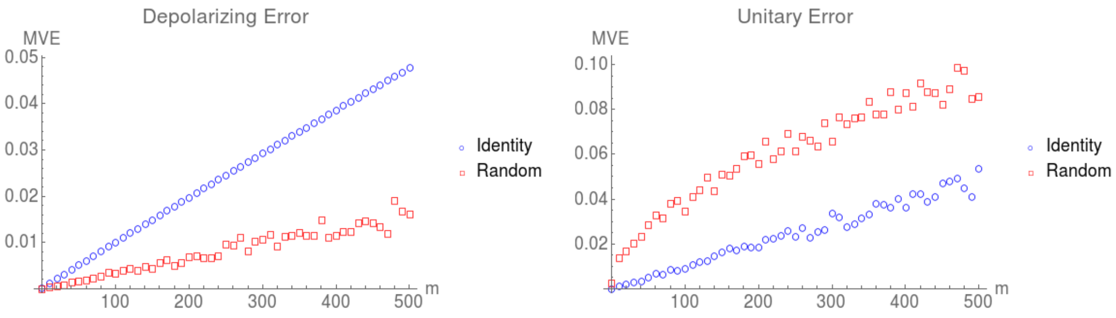

Figure 4.2: Mean variation error from Equation 4.18 for a gate-set with $\rho = M_1 = |0\rangle\langle 0|$, and $\mathcal{G} = \mathrm{Cl}_1$ being the 1-qubit Clifford group. Blue circles indicate self-inverting (identity) circuits whereas red squares indicate random circuits. Each point consists of averaging 200 random circuits with length m. The error channel on every gate has an averaged gate infidelity of 10^{-4}. The depolarizing error channel is $\mathcal{E}_D(\rho, r) = (1-2r)\rho + rI$, whereas the unitary error is $\mathcal{E}_U(\rho,\theta) = e^{-i\theta Z}$ with $\theta = \arccos\left(\sqrt{1-3r/2}\right)$. MVE may have different behaviors under different error types (m or $\sqrt{m}$) for a random circuit, as compared to the linear behavior for an identity circuit

defined by

$$\delta d(\vec{C}_m, \tilde{\vec{C}}_m) := \frac{1}{2}\sum_d \left|\langle \tilde{M}_d, \tilde{G}_{m:1}(\tilde{\rho})\rangle - \langle M_d, G_{m:1}(\rho)\rangle\right| \tag{4.16}$$

where the tilde represents real versions of the ideal operations, and we used the shorthand notation

$$G_{b:a} := \begin{cases} G_b G_{b-1} \ldots G_a & \text{if } b \geq a \\ \mathbb{1} & \text{otherwise} \end{cases} \tag{4.17}$$

to represent the series of gates being applied sequentially. This definition satisfies the requirements for being a metric (Definition 2.5.3), and is only based on measurable quantities. Denoting the set of all experiments with gate sequence length m as $\vec{\mathcal{C}}_m$, we further define the *Mean Variation Error* (MVE) over $\vec{\mathcal{C}}_m$ with the underlying gate-set $\mathbb{G}$ as

$$x(\mathbb{G}, m) := \mathbb{E}_{\vec{\mathcal{C}}_m}[\delta d(\vec{C}_m, \tilde{\vec{C}}_m)] \tag{4.18}$$

where $\mathbb{E}$ stands for averaging over a set. This measure was proposed in [82] as a "computationally relevant error per gate cycle", but the gauge-invariance property was not emphasized there. Note that although in each $\vec{C}_m$ only one state and one measurement is allowed (in order for the final outcome to be a valid probability distribution), there is no constraint on how many are included in the gate-set. The size of $\vec{\mathcal{C}}_m$ is thus given by $\left|\vec{\mathcal{C}}_m\right| = |\mathcal{S}||\mathcal{G}|^m|\mathcal{M}|$.

The MVE aims to quantify how well the apparatus performs all possible operations available from a target gate-set. In the special case where the measurement is a 2-outcome POVM and the gate sequence is signal-reversing (i.e. ideally an identity circuit), δd can be simplified as

$$\begin{aligned} \delta d(\vec{C}_m, \mathbb{1}) &= \frac{1}{2}\left(\left|\mathrm{Tr}\left[\tilde{M}_0^\dagger \tilde{G}_{m:1}(\tilde{\rho})\right] - 1\right| + \left|\mathrm{Tr}\left[(I - \tilde{M}_0)^\dagger \tilde{G}_{m:1}(\tilde{\rho})\right] - 0\right|\right) \\ &= 1 - \mathrm{Tr}\left[\tilde{M}_0, \tilde{G}_{m:1}(\tilde{\rho})\right] \end{aligned} \tag{4.19}$$

whose average over $\vec{\mathcal{C}}_m$ is nothing but 1 minus the "survival probability" plotted in a conventional randomized benchmarking experiment. In particular, one expects that for Markovian noise, the inverted decay curve can always be well approximated by a linear relation to first order in the average error rate [72, 61].

We simulated random circuits of varying length m sampled from the gate-set $\{\mathcal{S} = |0\rangle\langle 0|, \mathcal{G} = Cl_1, \mathcal{M} = |0\rangle\langle 0|$ (with Cl_1 denoting the 1-qubit Clifford group), where erroneous gates are represented as $\tilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$ with $\mathcal{E}$ being an error channel. We simulated two types of random circuits: circuits from the entire set of possible experiments allowed by the gate-set, and circuits from the set of possible experiments where the overall gate applied is identity (i.e., $G_{m:1} = \mathbb{1}$). In both simulations, the state and measurements are assumed to be error-free. The results are shown in Figure 4.2. When the error follows a depolarizing noise model, the MVE scales linearly with the gate sequence length m for both random and identity circuits, with the decreased rate for random circuits being $\sim 1/3$ the rate for identity circuits. This is because because when the state is transformed onto the xy-plane of the Bloch sphere right before measurement (which happens about 2/3 of the time), the depolarizing channel does not affect the outcome probability of a Z-axis measurement, resulting in an MVE of 0 for those circuit sequences. In contrast, when the error is a gate-independent unitary error, the scaling remains linear for the identity circuits but follows a $\sqrt{m}$ dependence on gate sequence length. According to arguments in [82], this is due to randomization of this error channel by the Clifford group operations, resulting in an accumulative average error rate that essentially behaves like a 1-d random walk in the step number m. The key result here is that the MVE behaves differently under different realistic error models for a general random circuit, which reveals additional information that are hidden by an identity circuit experiment like in randomized benchmarking.

Unlike other distance measures where an improvement in quality can be caused by a bias in choosing a gauge-representation, a decrease in MVE is unequivocally an improvement due to its gauge-invariance and because, by definition, the output probability distribution gets closer to ideal. An experimentalist can perform a feedback loop whereby they update the control parameters, rerun the MVE evaluation experiment and compare to the previous result to see if the error has decreased. Protocols that use feedback from experimental outcomes to improve control over quantum devices have been proposed before, such as in [84] where control parameters were optimized by minimizing the randomized benchmarking decay rate. While having the same advantages as [84] such as being gauge-invariant and robust to environmental noise, using MVE instead of the randomized benchmarking decay rate as the error measure can provide more insight into the nature of the error process.

A protocol for calibrating a QIP device is presented as follows:

1. Choose a gate-set $\mathbb{G} = \{\mathcal{S}, \mathcal{G}, \mathcal{M}\}$ of experimental interest.
2. Select N_G random experiments $\vec{C}_m$ with gate sequence length m, with N_G being large enough to accurately approximate $\mathbb{E}_{\vec{C}_m}$.
3. Compute the ideal outcome probabilities $\langle M_d, G_{m:1}(\rho)\rangle$ for each $\vec{C}_m$.
4. Repeatedly collect the measurement outcomes for N_M times to estimate $\langle \tilde{M}_d, \tilde{G}_{m:1}(\tilde{\rho})\rangle$ for each $\vec{C}_m$.
5. Calculate $\delta d(\vec{C}_m, \vec{\tilde{C}}_m)$ for each chosen $\vec{C}_m$, average over them to estimate $x(\mathbb{G}, m)$.
6. Repeat step 2-5 for different values of m to measure the scaling behaviour of the MVE.
7. Repeat steps 2-6 for different experimental control parameters, and track which parameter changes lower the MVE across all tested sequence lengths.

We briefly discuss the scalability of estimating the MVE here. In steps 2 and 4 in the above protocol, the magnitudes of N_G and N_M essentially determine the accuracy of counting experiments, and can be estimated using Hoeffding's bound [83] if a particular precision ϵ is desired with confidence $1 - \delta$. On the other hand, the scaling behavior of step 3 depends on the gates in $\mathcal{G}$: while this step can be efficiently computed if it only contains Clifford group elements according to the Gottesman-Knill theorem (Theorem 3.3.1), it is computationally hard if $\mathcal{G}$ also contains general unitary gates. Of course, being hard does not imply being impossible: for small systems with a few qubits, this procedure can still be performed quickly on a classical computer, which would still provide valuable information on the actual performance of a device implementing a particular gate-set. Further analysis on the behavior of this measure under different error models will be a highly desired result.

Chapter 5

Characterizing state and measurement operations

We have seen in the previous chapter how the gauge degree of freedom affects our description of a quantum system, and in particular how it prevents a fully self-consistent characterization in the absence of any assumptions. In other words, no experiment is possible to simultaneously determine every element in a gate-set without assuming the form of some elements. It is a question worth discussing that which of these assumptions are more credible than others. Many recent experimental efforts chose to characterize the quality of quantum gates by assuming ideal SPAM: this is not surprising considering that performing gates involves fine-tuning many parameters such as magnetic field strength and sequence duration, and is thus deemed more difficult than state preparation and measurements. Since the number of gates required in a certain computation task is generally quite large, it is reasonable to first improve gate performance at the earlier stage of quantum computing research.

On the other hand, one must be aware that the results from benchmarking quantum gates cannot be trusted if the SPAM being used to probe these gates are inherently faulty: this issue is becoming particularly non-negligible in some implementations. For example, in trapped ion systems, it is now possible to achieve an average gate fidelity on the order of 99% (and even 99.9% [85])for two-qubit gates and 99.99% for 1-qubit gates, whereas state and measurement infidelity is still on the order of 10^{-2} to 10^{-1} [8]. While these numbers are not directly comparable because they describe different processes, it is reasonable to anticipate SPAM errors to at least contribute on a similar scale as gate errors in these systems. As the quality of gates keep improving, the imperfectness on state preparation and measurement may start to become a bottleneck both in characterizing quantum gates and in the overall control precision. Moreover, many recent fault-tolerant schemes of quantum computation requires repeated rounds of state preparation and measurement operations, such as the surface code. This implies that SPAM errors are in fact just as important to overcome as gate errors. Therefore, protocols which focus on characterization of SPAM errors will be important in the near future where precise control on quantum devices are needed.

In this chapter, we introduce a method that characterizes single-qubit state and measurement errors separately and unambiguously, by assuming close-to-ideal quantum gates. The method uses common unitary gates and can be implemented on all known QIP platforms. In Section 5.1 we illustrate the idea behind this method, why it overcomes the aforementioned gauge ambiguity, and give the basic version of the protocol. In Section 5.2 we derive a bound on the effects of noisy quantum gates on the final output state in terms of the worst case error rate of multiqubit gates, then integrate a technique called randomized compiling to reduce the effect of the gate errors. Finally in Section 5.3, we give an approximate bound for the range of estimated parameters based on bounds on the output states, and simulated the experiment against an adversarial error model to demonstrate the robustness of our protocol.

5.1 Motivation and Basic Protocol

As we have seen previously, the gauge DoF effectively prevents separately characterizing the state and measurement errors: for example, in a direct measurement experiment without applying any external gates, one can assign the following gauge transformation

$$\langle\!\langle E||\rho\rangle\!\rangle = (\langle\!\langle E|B^{-1})(B|\rho\rangle\!\rangle) \tag{5.1}$$

which effectively includes all possibilities to assign the observed deviation from ideal to different models of SPAM errors. In particular, consider the following two gauge matrices, where the subscript t implies the desired (target) operation:

$$B_1 = |\rho_t\rangle\!\rangle\langle\!\langle\rho|,\ \ B_2 = |E\rangle\!\rangle\langle\!\langle E_t| \tag{5.2}$$

where the first will assign all the observed error to the measurement and outputs a perfect initial state, while the second does the reverse. Consequently it is impossible to talk about the individual errors on SPAM based only on this information. To achieve this goal, we must break this gauge symmetry between SPAM.

Our protocol is based on the following observation: if multiple state preparation is performed, and if we assume that the intrinsic error occurred during each preparation procedure is identical, then we can use multi-qubit gates to propagate these individual state preparation errors such that their effects can be measured on a single qubit. The final (potentially erroneous) measurement outcome from this qubit would have a contribution from combined state error that depends on the number of state preparations involved, and another fixed contribution from the measurement. The state and measurement parameters can then be extracted from the outcome pattern based on their different behaviors.

Let's first make this idea more concrete by considering the problem of characterizing a single qubit. We assume that the target state is $\rho_t = |0\rangle\langle 0|$, and we try to perform a projective two-outcome measurement $\mathcal{M}$ with operators $M_{1,t} = |0\rangle\langle 0|,\ M_{2,t} = |1\rangle\langle 1|$. In reality, the prepared state has a probability ϵ_1 to be in the other Z-eigenstate $|1\rangle\langle 1|$, allowing it to be described by a mixed state $\rho_r = \frac{1}{2}(I + (1-\epsilon_1)Z)$. The non-ideal measurement also has a probability ϵ_2 to output the opposite result given any Z-state inputs, and can be described by $M_{1,r} = \frac{1}{2}(I + (1-\epsilon_2)Z),\ M_{2,r} = \frac{1}{2}(I - (1-\epsilon_2)Z)$. Now consider the following circuit with inputs and measurements given as above. Since X-errors (an "impurity" in the Z-component) propagate from the control to the target bit through the CNOT gate, the bit-flip error happening during the state preparation will accumulate on the first bit as we increase the number of ancillary qubits. In the absence of gate errors, the "0" outcome probability is given by

$$pr_0(m) = \frac{1}{2}(1 + (1-\epsilon_1)^{m+1}(1-\epsilon_2)) \tag{5.3}$$

when a total of m gates are performed.

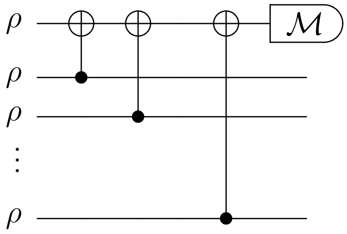

Figure 5.1: A state error amplifying circuit that transfers X-errors in state preparation to the first qubit.

By recording the results from a series of experiments with different m and fitting to the model $pr_0(m) = \frac{1}{2}(1 + ba^{m+1})$ to find the parameters a and b, one can obtain an estimate on both ϵ_1 and ϵ_2. The total number of qubits required for this method is not as large as suggested by the circuit: if state preparation (and reset) can be made fast enough compared to gate time: in such cases, the we can replace the previous circuit by the two-qubit one in Figure 5.2, where the reset step is simply resetting the second qubit using the same method as in preparing the state initially. Because a reasonable fitting for Equation 5.3 can be made with only a few data points (as there are only two degrees of freedom in the model), the state preparation time does not need to be indefinitely short compared to the gate time, so in principle as few as two qubits would be sufficient to extract the SPAM parameters. While in the following texts we still write the circuits in full for the sake of clarity, reduction to this simplified circuit is always possible for our presented protocols.

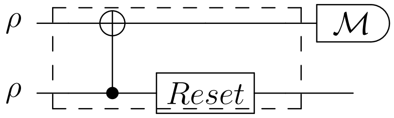

Figure 5.2: Simplified circuit for SPAM tomography method. Repeating the operations in the dashed box m times has the same effect as in the full circuit, Figure 5.1.

This simple method eliminates the possible gauge ambiguity for the following reason. The process of attaching an ancillary qubit in the state $\frac{1}{2}(I + (1 - \epsilon_1 Z))$, performing an ideal CNOT gate, and tracing out the ancilla can be expressed in Liouville form as

$$G_\Phi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1-\epsilon_1 & 0 \\ 0 & 0 & 0 & 1-\epsilon_1 \end{pmatrix} \tag{5.4}$$

The outcome probability for measuring the final state is $pr_0 = \langle\langle M_1 | G_\Phi^m | \rho \rangle\rangle$, which is a non-linear function in ϵ_1 for $m \geq 1$. Consequently, the transformation in Equation 5.1 fails to be a "gauge transformation" because it cannot, in general, preserve the outcome probability by linearly transforming the state parameters. This shows that in principle, state and measurement parameters can be determined individually, with the help of multi-qubit gates.

The above simplified analysis has two key limitations. First, instead of having only Z-components, the initial state and measurement are in fact described by a more general model

$$\rho = \frac{1}{2}(I + aX + bY + cZ),\ M_1 = \frac{1}{2}(dI + eX + fY + gZ) \tag{5.5}$$

with constraints on the parameters a-g such that ρ, M_1 and M_2 all satisfy the canonical constraints defined at the beginning of Chapter 2.3. Analyzing the same circuit with these input/outputs give the following probability for having a 0-outcome: $pr_{0,real}(m) = \frac{1}{2}(d + ae + c^m(bf + cg))$. This multi-variable expression hinders the simplicity of Equation 5.3 and would require more than a line-fit to solve individually. Of course, in the experimentally-relevant situation where both ρ and $\mathcal{M}$ are close to ideal, the parameters c, d, g will be close to 1 whereas the others will be close to 0, and ignoring second-order terms in these small parameters in $pr_{0,real}(m)$ will (almost) recover Equation 5.3 besides the extra factor d, allowing us to solve for c and g approximately. It may then be possible to plug these back in $pr_{0,real}(m)$ and iteratively solve for all parameters including the smaller ones. However, this would add complication to data and error analysis. Second, the quantum gates being applied in a real experiment would not be ideal. The error associated with each gate will reveal itself in the final outcome, but in a way unknown to us; and since we assume complete ignorance about states/measurements, we may not directly perform process tomography to learn about the exact form of the gates. Therefore, it is crucial that our protocol can bound the estimated parameters, with quality metrics on the gates that can be obtained even without knowing the SPAM parameters.

In the remaining part of this chapter and the next, we introduce a modified version of the protocol that resolves these problems. The important concept of randomization will be used in overcoming both issues. The first issue can be addressed by using a state randomization protocol: for any set $\mathcal{G}$ of unitary operators in $\mathcal{U}(\mathcal{H}^{2^n})$, we define the state randomization map $\mathcal{R}_0$ that applies a random unitary gate from $\mathcal{G}$ as

$$\mathcal{R}_0(\rho, \mathcal{G}) \coloneqq \frac{1}{|\mathcal{G}|} \sum_{\sigma_i \in \mathcal{G}} \sigma_i \rho \sigma_i^\dagger \tag{5.6}$$

The expectation (unbiased average) $\mathbb{E}$ of any linear function f of the state ρ after this randomization is then given by

$$\mathbb{E}_{\mathcal{G}}(f(\rho)) = \frac{1}{|\mathcal{G}|} \sum_{\sigma \in \mathcal{G}} f(\sigma^\dagger \rho \sigma) \tag{5.7}$$

For example, if we choose $\mathcal{G}$ to be $\mathcal{G} = \{I, X, Y, Z\}$, then randomization will result in an effective average state that is completely mixed; measurement on this state will yield an outcome 0 with probability

$$pr_0 = \mathrm{Tr}[M_1(I/2)] = d/2 \tag{5.8}$$

allowing us to determine the value of d. This can be shown by noting that conjugating one Pauli matrix by a different one will introduce an additional minus sign on the original one, i.e.,

$$P_j^\dagger P_i P_j = (-1)^{1-\delta_{ij}} P_i \quad \forall i,j \in \{x,y,z\} \tag{5.9}$$

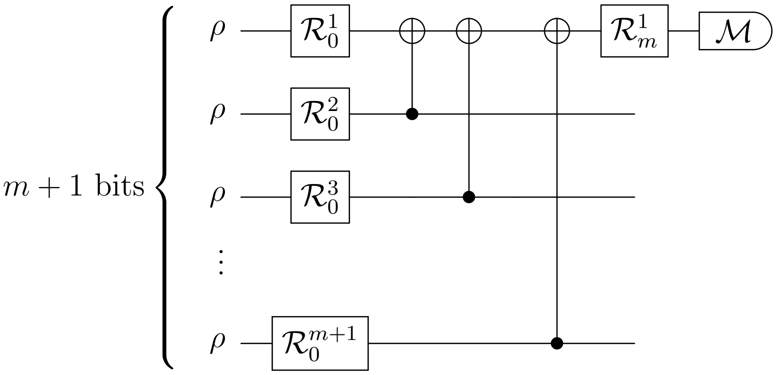

Figure 5.3: A modified state error amplifying circuit with initial state and measurement randomization. Each $\mathcal{R}_0^i$ represents a randomly chosen gate (independent from qubit to qubit) from a pre-defined set $\mathcal{G}$ in every run, and experiments with different random gates are performed many times to obtain the expectation value over $\mathcal{G}$ to a desired precision.

It is then easy to see that, if in a different experiment we set $\mathcal{G} = \{\mathbb{1}, Z\}$, the averaged output state will have I and Z components only. And since Pauli channels are their own adjoint, performing the same randomization with $\mathcal{G} = \{\mathbb{1}, Z\}$ right before the measurement will turn the measurement operator into having only I and Z components too. Repeating the same circuit as in Figure 5.1 gives an expression for measuring a 0-outcome as

$$\begin{aligned} pr_0(m) &= \frac{1}{2}(d + c^{m+1}g) \\ 2pr_0(m) - d &= c^{m+1}g \end{aligned} \tag{5.10}$$

which allows us to determine c and g in a line fit with the previous knowledge about d.

Once the Z component of the state and measurement operators are determined, one can in principle estimate all other components by performing single-qubit rotations (in the Bloch-sphere picture, Figure 2.1) and using the Z-states as a known "probe", just as in standard state (measurement) tomography experiments with a known but imperfect detector (probe state), using the reconstruction methods introduced in Section 3.2. For example, to measure the component a, one could first perform state randomization with $\mathcal{G} = \{\mathbb{1}, X\}$, followed by a $Y_{-\pi/2}$ pulse that transforms it into a Z-state, and measure with the same measurement. Note that it would be both unwise and unnecessary to repeat circuit 5.1 just to probe these smaller components: if $\epsilon_1 \approx 1$ in Equation 5.3, $pr_0(m)$ will quickly become indistinguishable from $\frac{1}{2}$ for relatively small values of m, resulting in a large relative error in the estimated parameters.

5.2 Reducing Gate Errors with Randomized Compiling

We now address the second problem mentioned previously and analyze the effect of gate errors. First, we derive a bound on the distance between final states in the presence and absence of gate errors, using the diamond norm of a perturbation term in the gate that represents its error; second, we apply a technique called randomized compiling [18] to modify the previous bound by replacing the diamond norm error with an average gate infidelity, which can be much smaller than the diamond norm error and may be estimated in the presence of SPAM errors.

To start, we perturbatively expand every linear CPTP map $\tilde{\mathcal{A}}$ as $\tilde{\mathcal{A}} = \mathcal{A}_t + \delta\mathcal{A}$, where $\delta\mathcal{A}$ can be considered as the gate error. For now we can consider gate-dependent errors so that $\delta\mathcal{A}_i$ is different for every different gate G_i. We will only assume that the noise is Markovian, so that they can be represented as a matrix in Liouville form. Since both $\mathcal{A}$ and $\tilde{\mathcal{A}}$ represent CPTP maps, $\delta\mathcal{A}$ must be a trace-eliminating map. Now, consider the state of qubit 1 in the circuit 5.3: with a total of m ancillary qubits, it can be written as

$$\begin{aligned}\tilde{\rho}_{rf} &= \mathrm{Tr}_{2\ldots m+1}\left[\left(\bigcirc_{n=2}^{m}\tilde{\mathcal{C}}_{n,1}\right)\circ\left(\bigotimes_{n=1}^{m+1}\tilde{\mathcal{R}}_{n,\tau}\right)(\rho_r)^{\otimes m+1}\right]\\ &= \mathrm{Tr}_{2\ldots m+1}\left[(\bigcirc_{n=2}^{m}(\mathcal{C}_{n,1}+\delta\mathcal{C}_{n,1}))\circ\left(\bigotimes_{n=1}^{m+1}(\mathcal{R}_{n,\tau}+\delta\mathcal{R}_{n,\tau})\right)(\rho_r)^{\otimes m+1}\right]\end{aligned} \tag{5.11}$$

where $\mathcal{R}_{n,\tau}$ stands for the random gate (selected as $\sigma_\tau \in \mathbf{P}_1$) applied on qubit n, and $\mathcal{C}_{a,b}$ is a CNOT gate that is controlled on qubit a and targeted on qubit b. Comparing with the final state in the absence of gate errors, ρ_{rf},

$$\rho_{rf} = \mathrm{Tr}_{2\ldots m+1}\left[(\bigcirc_{n=2}^{m}\mathcal{C}_{n,1})\circ\left(\bigotimes_{n=1}^{m+1}\mathcal{R}_{n,\tau}\right)(\rho_r)^{\otimes m+1}\right] \tag{5.12}$$

the first order term in the difference between these two gates is given by

$$\begin{aligned}\rho_{rf} - \tilde{\rho}_{rf} = \mathrm{Tr}_{2\ldots m+1}\Bigg[&\left(\sum_{n'=2}^{m+1} \bigcirc_{n=2}^{m+1}((1-\delta_{n,n'})\mathcal{C}_{n,1} + \delta_{n,n'}\delta\mathcal{C}_{n,1})\right) \circ \left(\bigotimes_{n=1}^{m+1} \mathcal{R}_{n,\tau}\right) + \\ &(\bigcirc_{n=2}^{m}\mathcal{C}_{n,1}) \circ \left(\sum_{n'=2}^{m+1}\bigotimes_{n=1}^{m+1}((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n,\tau})\right)\Bigg](\tilde{\rho}) + \mathcal{O}(\delta^2) \end{aligned} \tag{5.13}$$

where $\mathcal{O}(\delta^2)$ includes all terms of second order or higher in either $\delta\mathcal{C}$ or $\delta\mathcal{R}$. The distance between these two states reflects the error introduced by the imperfect gates. We quantify this distance by bounding the 1-norm of the above first order term.

First note that by triangle inequality, the two sums in the above expression (which we call $\mathbb{A}$ and $\mathbb{B}$) can be separately bounded:

$$\|\rho_{rf} - \tilde{\rho}_{rf}\|_1 \leq \|\mathbb{A}\|_1 + \|\mathbb{B}\|_1 + \mathcal{O}(\delta^2) \tag{5.14}$$

where

$$\begin{aligned}\mathbb{A} &= \mathrm{Tr}_{2\ldots m+1}\left[\left(\sum_{n'=2}^{m+1} \bigcirc_{n=2}^{m+1}((1-\delta_{n,n'})\mathcal{C}_{n,1} + \delta_{n,n'}\delta\mathcal{C}_{n,1})\right) \circ \left(\bigotimes_{n=1}^{m+1} \mathcal{R}_{n,\tau}\right)\right](\tilde{\rho}) \\ \mathbb{B} &= \mathrm{Tr}_{2\ldots m+1}\left[(\bigcirc_{n=2}^{m}\mathcal{C}_{n,1}) \circ \left(\sum_{n'=1}^{m+1}\bigotimes_{n=1}^{m+1}((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n,\tau})\right)\right](\tilde{\rho}) \end{aligned} \tag{5.15}$$

Both terms can be bounded in a similar manner: first,

$$\begin{aligned}
\|\mathbb{A}\|_1 &\leq \sum_{n'=2}^{m+1} \left\| \mathrm{Tr}_{2\ldots m+1} \left(\bigcirc_{n=2}^{m+1} ((1-\delta_{n,n'})\mathcal{C}_{n,1} + \delta_{n,n'}\delta\mathcal{C}_{n,1}) \right) \circ \left(\bigotimes_{n=1}^{m+1} \mathcal{R}_{n,\tau} \right) (\tilde{\rho}) \right\|_1 \\
&\leq \sum_{n'=2}^{m+1} \left\| \left(\bigcirc_{n=2}^{m+1} ((1-\delta_{n,n'})\mathcal{C}_{n,1} + \delta_{n,n'}\delta\mathcal{C}_{n,1}) \right) \circ \left(\bigotimes_{n=1}^{m+1} \mathcal{R}_{n,\tau} \right) (\tilde{\rho}) \right\|_1 \\
&\leq \sum_{n'=2}^{m+1} \left\| \left(\bigcirc_{n=2}^{m+1} ((1-\delta_{n,n'})\mathcal{C}_{n,1} + \delta_{n,n'}\delta\mathcal{C}_{n,1}) \right) \right\|_\diamond \\
&\leq \sum_{n'=2}^{m+1} \prod_{n}^{m+1} \|((1-\delta_{n,n'})\mathcal{C}_{n,1} + \delta_{n,n'}\delta\mathcal{C}_{n,1})\|_\diamond \\
&= \sum_{n=2}^{m+1} \|\delta\mathcal{C}_{n,1}\|_\diamond
\end{aligned}$$

where from the top to bottom we used the triangle inequality, monotonicity of 1-norm with respect to partial trace (Lemma 2.5.1), definition of the diamond norm(Equation 2.49), submultiplicativity of the diamond norm (Equation 2.53), and the fact that $\|\Phi\|_\diamond = 1$ for every CPTP map Φ.

Similarly,

$$\begin{aligned}
\|\mathbb{B}\|_1 &\leq \sum_{n'=1}^{m+1} \left\| \mathrm{Tr}_{2\ldots m+1} \left[\left(\bigcirc_{n=2}^{m} \mathcal{C}_{n,1} \right) \circ \left(\bigotimes_{n=1}^{m+1} ((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n,\tau}) \right) \right] (\tilde{\rho}) \right\|_1 \\
&\leq \sum_{n'=1}^{m+1} \left\| \left[\left(\bigcirc_{n=2}^{m} \mathcal{C}_{n,1} \right) \circ \left(\bigotimes_{n=1}^{m+1} ((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n,\tau}) \right) \right] \right\|_\diamond \\
&\leq \sum_{n'=1}^{m+1} \left(\prod_{n=2}^{m+1} \left\| \bigcirc_{n=2}^{m} \mathcal{C}_{n,1} \right\|_\diamond \right) \left\| \bigotimes_{n=1}^{N+1} ((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n,\tau}) \right\|_\diamond \\
&= \sum_{n'=1}^{m+1} \left\| \bigotimes_{n=1}^{m+1} ((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n,\tau}) \right\|_\diamond \\
&= \sum_{n'=1}^{m+1} \prod_{n=1}^{m+1} \left\| ((1-\delta_{n,n'})\mathcal{R}_{n,\tau} + \delta_{n,n'}\delta\mathcal{R}_{n',\tau}) \right\|_\diamond \\
&\leq \sum_{n'=1}^{m+1} \|\delta\mathcal{R}_{n,\tau}\|_\diamond
\end{aligned}$$

where from top to bottom we used the triangle inequality, monotonicity of 1-norm and definition of diamond norm, submultiplicativity of diamond norm, $\|\Phi\|_\diamond = 1$ for a CPTP map Φ, and multiplicativity of diamond norm with respect to tensor product of channels (Equation 2.54). Summarizing, we have

$$\|\rho_{rf} - \tilde{\rho}_{rf}\|_1 \lesssim \sum_{n'=2}^{m+1} \|\delta\mathcal{C}_{n,1}\|_\diamond + \sum_{n=1}^{m+1} \|\delta\mathcal{R}_{n,\tau}\|_\diamond \tag{5.16}$$

Separating a map into the ideal plus a deviation map allows us to obtain this first-order bound. On the other hand, the worst-case error (2.56) are defined for an error *channel*, while the deviation map is not a channel since it is trace-eliminating. In order to relate terms in Equation 5.16 to the worst-case error, note that for every real map $\tilde{G}$ decomposed as $G + \delta G$ where G is an ideal unitary channel, we can always write it equivalently as $G \circ \mathcal{E}$

for some noise channel $\mathcal{E}$ (see Figure 5.4). Then

$$\begin{aligned} G \circ \mathcal{E} &= G + \delta G \\ \mathcal{E} &= \mathbb{1}_{T(\mathcal{H}^n)} + G^\dagger \circ \delta G \\ \mathcal{E} - \mathbb{1}_{T(\mathcal{H}^n)} &= G^\dagger \circ \delta G \\ \left\| \mathcal{E} - \mathbb{1}_{L(\mathcal{H}^n)} \right\|_\diamond &= \left\| G^\dagger \circ \delta G \right\|_\diamond = \|\delta G\|_\diamond \end{aligned}$$

where the last line follows from the unitary invariance of $\|\cdot\|_{1\to 1}$ (Equation 2.50 and 2.31) and the fact that $G^\dagger(X) = U^\dagger X U$ for some unitary operator U. This relation holds as long as the target channel is unitary. Therefore, we may rewrite Equation 5.16 with more commonly used quantities as

$$\|\rho_{rf} - \tilde{\rho}_{rf}\|_1 \lesssim \sum_{n=2}^{m+1} \epsilon(\mathcal{E}(\mathcal{C}_{n,1})) + \sum_{n=1}^{m+1} \epsilon(\mathcal{E}(\mathcal{R}_{n,\tau})) \tag{5.17}$$

with a different interpretation of gate error, where $\epsilon(\mathcal{E})$ is the worst case error rate for a noise channel $\mathcal{E}$ defined in Equation 2.56.

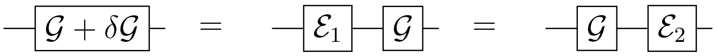

Figure 5.4: Different ways to represent gate errors.

What we have essentially shown is that to first order in the error channels, the effect of these erroneous gates on the final output state is upper bounded by the sum of diamond norms on all the pre-multiplied noisy channels on all the gates in circuit 5.3. However, the diamond norm of these deviation maps are not experimentally accessible unless full process tomography is performed on every gate. Most experimental characterizations of quantum gates report their quality in terms of the estimated average gate infidelity $r(\mathcal{E})$ for a noise channel $\mathcal{E}$, which is related to $\epsilon(\mathcal{E})$ via the following bounds [86, 57]

$$r(\mathcal{E})\frac{d+1}{d} \leq \epsilon(\mathcal{E}) \leq \sqrt{r(\mathcal{E})}\sqrt{d(d+1)} \tag{5.18}$$

It is known that the lower bound is saturated by Pauli noise channels[1], whereas the upper bound is saturated by pure unitary noise, up to some constant factors proportional

---

[1] A Pauli channel has a Kraus representation (Equation 2.10) with all Kraus operators being (tensor product of) Pauli operators.

to $\sqrt{d}$ [11]. The upper bound scaling is clearly suboptimal in terms of the infidelity because of the square root: it has been shown that one can easily construct a noise channel $\mathcal{E}$ for a two-qubit generalized controlled phase gate with $r(\mathcal{E}) = 0.01$, but having a worst case error rate of $\epsilon(\mathcal{E}) \approx 0.13$; in fact, the upper bound for arbitrary two-qubit noise channels with $r(\mathcal{E}) = 0.01$ is as high as 0.45 [11]. Therefore, the RHS of Equation 5.17 cannot be deemed small for an arbitrary noise channel unless it has special properties. Our solution to this problem is to modify our protocol using a technique called *randomized compiling* (RC) invented by Wallman and Emerson [18]. The idea of RC is to insert error-randomizing gates between the "hard-gates" (in our case the CNOT gate) in order to tailor the noise on these gates to be a stochastic Pauli noise, which saturates the lower bound in Equation 5.18. Since the average infidelity of each gate can be estimated independent of SPAM errors, our method has the advantage that both the estimated parameters and the bound on them are obtained without prior assumptions about either the states or measurements, in contrast to conventional state or measurement tomography.

The steps to integrate RC technique into our protocol are summarized as follows. First, we insert a virtual identity gate $\mathbb{1}_{\mathcal{H}^{2^{m+1}}}$ between each neighboring CNOT gates so that the protocol now have exactly the same form as the setup in [18]: that the circuit becomes a sequence of alternating "simple" and "hard" gates corresponding to single-qubit Pauli gates and two-qubit CNOT gates in our setup. Next, we insert a single-qubit gate T_i^j right after each original single-qubit gate, and a correction gate T_i^c right before the next series of single qubit gates, including the virtual identity gates (see Figure 5.5 (b)). T_i^j is chosen randomly and independently from the set $\mathbf{P}_1$ at each position and in each run. The correction gates are uniquely determined by the previous single-qubit randomizing gates, defined as

$$\vec{T}_i^c := \mathcal{C}_{i,1}\vec{T}_i^\dagger(\mathcal{C}_{i,1})^\dagger \ \forall i \in \{0,..,m-1\} \tag{5.19}$$

so that the circuit after applying $\vec{T}_i^c\mathcal{C}_{i,1}\vec{T}_i$ is logically equivalent to the original circuit [2] (we used the vector notation $\vec{T}_i$ to denote the tensor-product channel on all $(m+1)$ qubits at step i). Finally, the intermediate gates are compiled into a single gate $\hat{C}_i$ according to the following rule

$$\begin{gathered}\hat{\vec{\mathcal{R}}}_0 := \vec{T}_0\vec{\mathcal{R}}_0,\ \hat{\mathcal{R}}_1 := \mathcal{R}_1\\ \hat{\vec{C}}_i := \vec{T}_i\vec{T}_{i-1}^c \ \forall i \in \{1,...,m-1\}\\ \hat{\vec{C}}_m := \vec{T}_{m-1}^c\end{gathered} \tag{5.20}$$

---

[2] Note that the notion $\mathcal{AB}$ for two quantum operations $\mathcal{A}$ and $\mathcal{B}$ implies an action on the operator X in the order $\mathcal{A}(\mathcal{B}(X))$.

Note that the compiled gates $\hat{\vec{C}}_i$ all belong to the set of $(m+1)$-fold tensor product of Pauli channels, so there is no need to introduce new gates in addition to single-qubit Pauli gates that are used to randomize the initial states. This is possible because the only hard gate in our setup (the CNOT gate) belongs to the Clifford group that maps Pauli operators to Pauli operators.

The formal protocol of the experiment is summarized as follows:

0. Measure the probability of obtaining outcome 0 (pr_0) for the qubit prepared in the target state $\rho_t = |0\rangle\langle 0|$ multiple times until desired precision is obtained, compute the parameter d using Equation 5.8;
1. Uniformly sample to generate a list of random gates $\mathcal{R}_0^j$, $\mathcal{R}_m^1 \in \{I, Z\}$ and $T_i^j \in \{I, X, Y, Z\}$ for $i = 0...m-1$ (denotes the i-th time step) and $j = 1...m+1$ (denotes the j-th qubit), and collect to form $\vec{\mathcal{R}}_0$ and $\vec{T}_i$;
2. Compile the gates according to Equation 5.20;
3. Implement the circuit (c) in Figure 5.5, collect the measurement outcome $x \in \{0, 1\}$;
4. Repeat steps 1-3 until a desired precision on the measurement outcome frequency $pr_0(m)$ is obtained;
5. Repeat steps 1-4 for $m = 1, 2, 3...m_{\max}$, fit to the decay model in Equation 5.10 and extract parameter c and g.
6. Perform the state randomization - single-qubit rotation - measurement protocol discussed at the end of Section 5.1 to extract all single-qubit parameters.

5.3 Bounding Parameters and Simulation

We now address the issues mentioned in Section 5.1 about bounding the precision of estimation and improving the worst case bound after randomized compiling. First we quantify the benefits from RC using the results in [18]. To reduce complications from gate-dependent noise in the simple gates, we further assume that all single-qubit Pauli gates can be performed perfectly, i.e., $\mathcal{E}(\mathcal{R}_{n,i}) = \mathcal{E}(T_{n,i}) = \mathbb{1}_{L(\mathcal{H}^2)}\ \forall n \in \{1...m+1\},\ i \in \{x, y, z\}$, and we only consider the major contribution of errors coming from the two-qubit CNOT gates.

Under this assumption, according to Theorem 1 in [18], the effective error channel on the CNOT gates will be tailored to $\mathcal{T}_j$ where

$$\mathcal{T}_j = \mathbb{E}_{\vec{T}}\vec{T}^{\dagger}\mathcal{E}(\mathcal{C}_{k,1})\vec{T} \tag{5.21}$$

which, according to [52], is a Pauli channel if the $\vec{T}$'s are sampled from Pauli channels. This holds true as long as the gate error $\mathcal{E}$'s are Markovian. Thus, it satisfies the lower bound in Equation 5.18. If we further assume that the error on CNOT gate is qubit-independent, i.e. $\mathcal{E}(\mathcal{C}_{i,1}) = \mathcal{E}(\mathcal{C})$, then we obtain from Equation 5.18 that

$$\begin{aligned} \|\rho_{rf} - \tilde{\rho}_{rf}\|_1 &\lesssim \sum_{n=2}^{m+1} 2\epsilon(\mathcal{E}(\mathcal{C}_{n,1})) \\ &= 2m\epsilon(\mathcal{E}(\mathcal{C})) \\ &= 3mr(\mathcal{E}(\mathcal{C})) \end{aligned} \tag{5.22}$$

Using this result together with Hölder's inequality (2.34), the difference between the final measured probabilities in the absence (pr_f) and presence ($\tilde{pr}_f$) of gate errors can be bounded as

$$\left|pr_f(m) - \tilde{pr}_f(m)\right| = |\langle M_1, \rho_{rf} - \tilde{\rho}_{rf}\rangle| \leq \|M_1\|_\infty \|\rho_{rf} - \tilde{\rho}_{rf}\|_1 \leq 3mr(\mathcal{E}(\mathcal{C})) \tag{5.23}$$

where the last inequality comes from the fact that $\|M_1\|_\infty = s_1(M_1)$ where $s_1(M_1)$ is its largest singular value [12], M_1 is positive-semidefinite (implying that $s_i(M_1) = \lambda_i(M_1)$ where $\lambda_i(M_1)$ are its eigenvalues) and $0 \leq M_1 \leq \mathbb{1}_{\mathcal{H}^2}$ (implying that its largest eigenvalue satisfies $\lambda_{\max}(M_1) \leq 1$). From Equation 5.10 and defining

$$y(m) := 2pr_f(m) - d, \tag{5.24}$$

a plot of y vs. m would ideally yield an exponentially-decaying curve gc^{m+1}; from Equation 5.23 it is clear that the value of y in the absence of gate errors is bounded to first order in $r(\mathcal{E}(\mathcal{C}))$ by

$$y(m) \in [\tilde{y}(m) - 6mr(\mathcal{E}(\mathcal{C})), \tilde{y}(m) + 6mr(\mathcal{E}(\mathcal{C}))] \tag{5.25}$$

where $\tilde{y}(m) = (2\tilde{pr}_f(m) - d)$. Thus, each data point recording the 0-outcome probability would have a bound in their value which grows linearly with m.

In order to obtain an estimated bound on the fitted parameters c and g from the bound on y, it is more straightforward to take the log on both sides of Equation 5.10 which linearizes the fitting model. This change of variable will alter the upper and lower bounds on y in Equation 5.25 into

$$\log(y) \in [\log(\tilde{y}) - \frac{6mr(\mathcal{E}(\mathcal{C}))}{\tilde{y}},\ \log(\tilde{y}) + \frac{6mr(\mathcal{E}(\mathcal{C}))}{\tilde{y}}] \tag{5.26}$$

to first order in r. Since the linearized fitted model is

$$\log(y(m)) = \log(g) + (m+1)\log(c), \tag{5.27}$$

we can obtain upper and lower bounds on $\log(c)$ by fitting the upper and lower bounds of $\log(y)$ (denoted as c_{up} and c_{low}) in Equation 5.26. By considering that when $m = 0$ there is no contribution from gate errors, we can approximately write

$$\log(y(m=0)) = \log(g_{\text{low}}) + \log(c_{\text{up}}) = \log(g_{\text{up}}) + \log(c_{\text{low}}) \tag{5.28}$$

which gives and approximate bound on $\log(g)$ and consequently g. More concretely, assuming that the lower/upper bounds on all parameters are not too far away from the fitted value, we have

$$\log(\tilde{y}) + \frac{6mr(\mathcal{E}(\mathcal{C}))}{\tilde{y}} \approx \log(g_{\text{fit}}) + (m+1)\log(c_{\text{up}}) \tag{5.29}$$

and

$$\log(\tilde{y}) - \frac{6mr(\mathcal{E}(\mathcal{C}))}{\tilde{y}} \approx \log(g_{\text{fit}}) + (m+1)\log(c_{\text{low}}) \tag{5.30}$$

where the subscript "fit" is used to denote the parameter value returned from fitting. Thus,

$$\begin{aligned} \frac{12mr(\mathcal{E}(\mathcal{C}))}{\tilde{y}} &\approx (m+1)\log\left(\frac{c_{\text{up}}}{c_{\text{low}}}\right) \\ c_{\text{up}} &\approx \exp\left(\frac{12r(\mathcal{E}(\mathcal{C}))}{\tilde{y}}\frac{m}{m+1}\right)c_{\text{low}} \\ c_{\text{up}} - c_{\text{low}} &\approx \frac{12r(\mathcal{E}(\mathcal{C}))}{\tilde{y}}\frac{m}{m+1} \approx 12r(\mathcal{E}(\mathcal{C})) \end{aligned} \tag{5.31}$$

And from Equation 5.28, it is easily seen that the difference between upper and lower bounds for g equals to that for c, so

$$g_{\text{up}} - g_{\text{low}} \approx 12r(\mathcal{E}(\mathcal{C})) \tag{5.32}$$

again to first order in r. The bound on the difference between fitted parameters and their true value may be estimated as half the difference between the upper and lower bounds.

We simulated the experiment with and without randomized compiling to compare the estimation accuracy. In our simulation, the state and measurement parameters are $c_{\text{real}} = g_{\text{real}} = d_{\text{real}} = 0.95$ and $a_{\text{real}} = b_{\text{real}} = e_{\text{real}} = f_{\text{real}} = 0.05$. All single-qubit Pauli gates are assumed to be ideal, while the real CNOT gate is modeled with a "small" unitary rotation gate $\mathcal{E}_{n,1} \coloneqq (X_1)^{\delta\theta} \otimes \mathbb{1}_n$ followed by an ideal CNOT gate $\mathcal{C}_{n,1}$, which systematically over-rotates the qubit being measured at the end. The matrix power $\delta\theta$ in $\mathcal{E}_{n,1}$ was chosen such that it has an average gate infidelity of $r(\delta\theta)$, calculated using the formula given in Theorem 2.5.3 [19]. For each value of m, we uniformly sample 200 random circuits from all possible ones with the same length to approximate the averaged output; and in each experiment we used $m_{\max} = 6$ ancillary qubits to perform fitting, based on realistic experimental considerations. We also ignored statistical error in the final outcome probability for each sequence. The results are shown in Figure 5.6: first, we show that for large values of r, the survival curve without randomized compiling no longer follows the model given by 5.27, whereas this effect is significantly reduced after compiling (see Figure 5.6 (iii) and (iv)). A large inconsistency between the fitted and real parameters is observed for almost all values of r in the non-compiled experiments [3], while the estimate accuracy are bounded by our estimated bound for both parameters c and g (the blue dashed curves corresponding to Equation 5.31 and 5.32) when randomized compiling is used. One may also notice that the approximate upper bound for a general noise model (red dotted curve, using the upper bound for $\epsilon(\mathcal{E})$ in Equation 5.18) is very loose, which supports our argument in Section 5.2 that a small average gate infidelity for a general error channel does not necessarily imply a small worst-case error rate, especially for small errors. Overall, we demonstrated that our method provides a good estimate for the SPAM parameters under a relatively strong adversarial error model.

---

[3] The exponential scaling is largely due to our reconstructing method, because the desired values are obtained from re-exponentiating the fitted parameters $\log(c_{\text{fit}})$ and $\log(g_{\text{fit}})$.

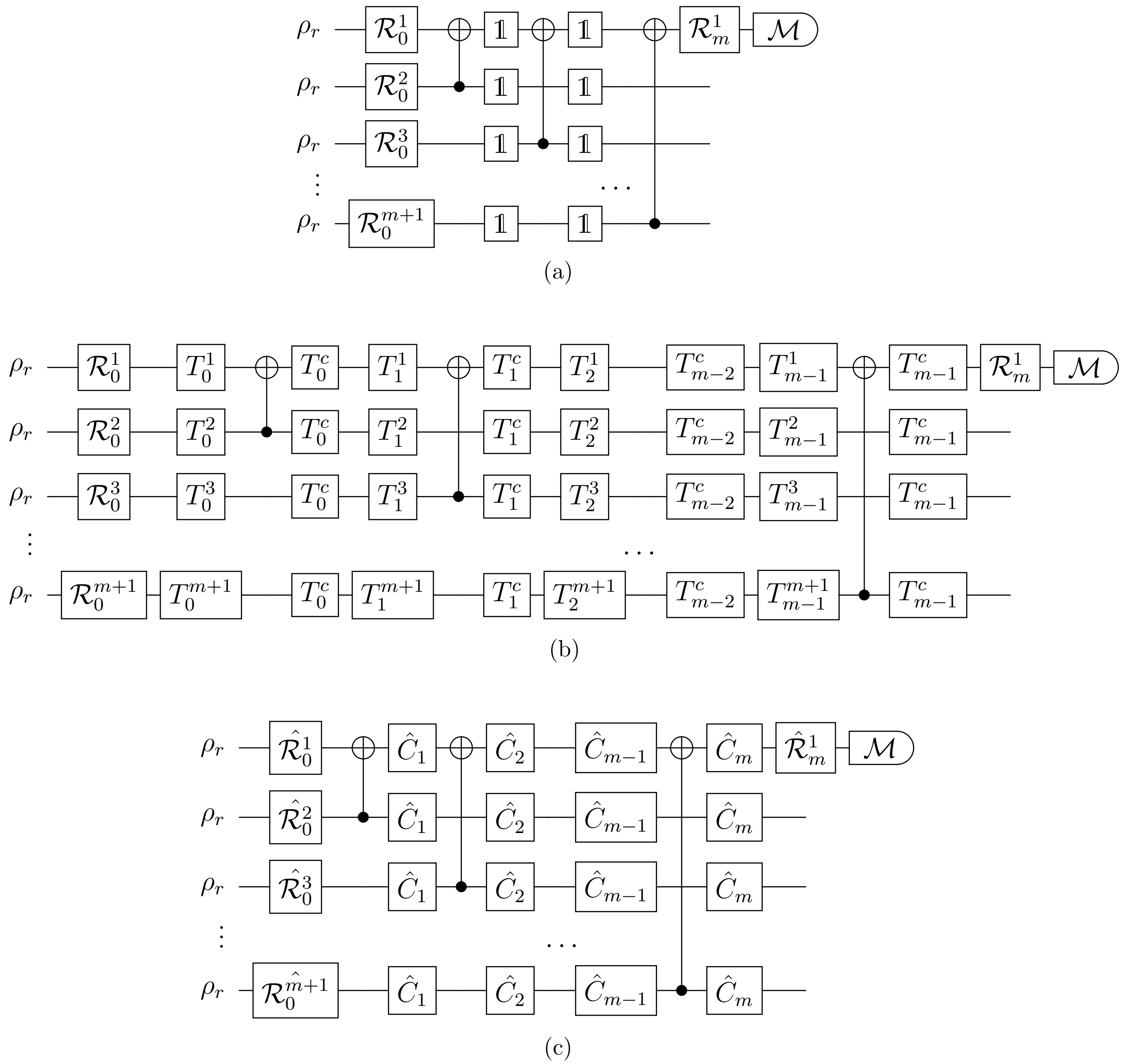

Figure 5.5: Example of state error detection circuit with state/measurement randomization and randomized compiling. (a) is the original circuit with virtual identity gates inserted. (b) contains error-randomizing single-qubit gates T_i^j, each being a random gate from the set $\mathbf{P}_1$ on qubit j; the correction gates are $\vec{T}_i^c = \mathcal{C}_{i,1}\vec{T}_i^\dagger(\mathcal{C}_{i,1})^\dagger$. (c) is the equivalent, compiled version of (b).

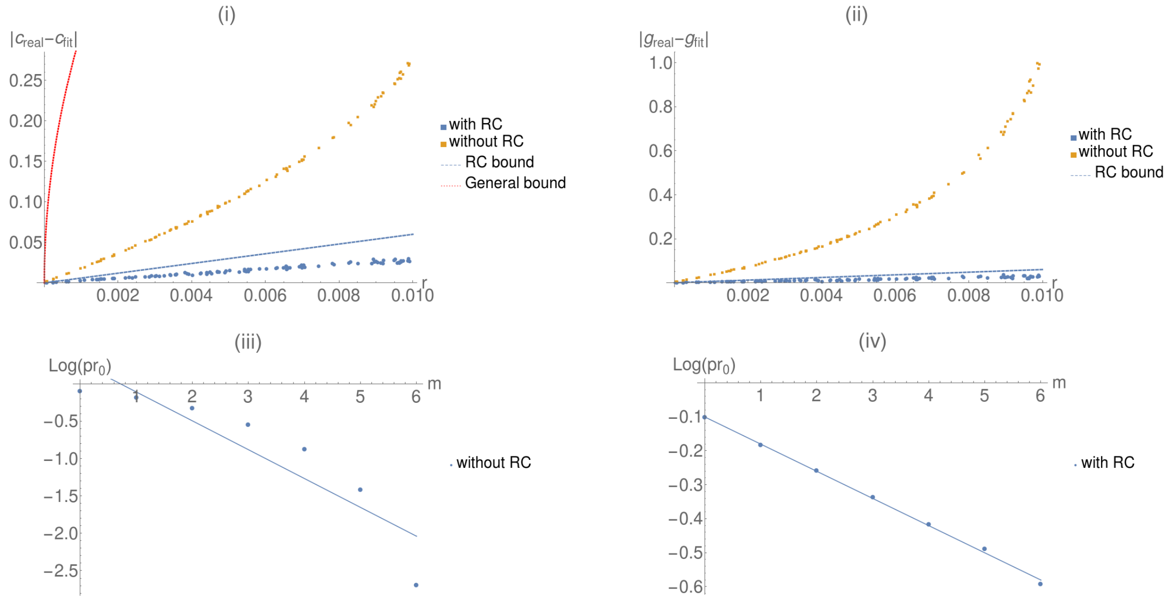

Figure 5.6: Numerical results of SPAM parameter estimation protocol. (i) - (ii): Plot of $|c_{\text{real}} - c_{\text{fit}}|$ (i) and $|g_{\text{real}} - g_{\text{fit}}|$ (ii) from simulated experiments, with the x-axis being the average gate infidelity for the imperfect CNOT gate. The blue dashed lines are the approximate upper bounds (Equation 5.31 and 5.32) for our protocol. Significant improvement in estimation accuracy and favorable scaling with $r(\mathcal{E})$ is observed with randomized compiling. (iii) - (iv): example of measured 0-state outcome frequency (Equation 5.10) as a function of m and a least-squares fitting line using Equation 5.27, with the same average infidelity $r(\mathcal{E}) = 0.01$ for the CNOT gate. (iii) and (iv) are implemented without and with RC, respectively. The undesired effect of this systematic error is significantly reduced by randomized compiling.

Chapter 6

Conclusion and Future Work

In this thesis, we have focused on two different aspects concerning how to correctly represent and characterize noisy QIP devices. In the first half, we demonstrated how a gauge DoF arises when representing a physical apparatus with a gate-set including a finite number of initial states, gates, and measurements, in the experimentally-relevant case where only measured probabilities are available. We discussed how conventional quality metrics for individual gate-set elements lost their operational meaning, when the object being characterized is determined from a tomography experiment. Finally, an alternative operational quality measure for a gate-set was proposed, and a method to improve experimental control that is not biased by the gauge was also introduced. In the second half, we propose a protocol which separately characterizes all single-qubit state and measurement parameters that is not biased by the gauge ambiguity, based on the assumption that high-quality quantum gates are available. We improved the method by integrating a technique called randomized compiling, and demonstrated its performance by simulating the protocol under an adversarial systematic unitary error model on the CNOT gate. Our approach provides approximate bounds on the estimated parameters based on the average gate infidelity, which can be efficiently estimated without knowing the SPAM parameters. This provides an alternative approach to benchmarking QIP devices that is, in some sense, complementary to conventional process tomography.

Several points that were left unexplored in this thesis may be considered as interesting future work. First, it would be useful to analyze the exact behavior of the gate-set mean variation error under different error models in the case of random circuits, which allows an experimentalist to infer quantitatively both the error type and strength based on the protocol introduced at the end of Chapter 4.4. For the SPAM characterization protocol introduced in Chapter 5, it is both interesting and useful to extend the protocol to multiqubit systems, which requires more sophisticated experimental design and post-analysis. It would be also very helpful to include a detailed error-analysis that provides a more realistic bound which depends on the number of random sequences used to approximate the population average, as well as the number of repeated measurements per sequence, thereby providing a fully instructive guide to real experiments.

References

[1] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[2] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Review, 41(2):303–332, 1999.

[3] Lov K Grover. A fast quantum mechanical algorithm for database search. In Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pages 212–219. ACM, 1996.

[4] Daniel R Simon. On the power of quantum computation. SIAM journal on computing, 26(5):1474–1483, 1997.

[5] Rami Barends, Julian Kelly, Anthony Megrant, Andrzej Veitia, Daniel Sank, Evan Jeffrey, Ted C White, Josh Mutus, Austin G Fowler, Brooks Campbell, Alexander N Korotkov, Andrew N Cleland, and John M Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. Nature, 508(7497):500, 2014.

[6] M Veldhorst, JCC Hwang, CH Yang, AW Leenstra, Bob de Ronde, JP Dehollain, JT Muhonen, FE Hudson, Kohei M Itoh, and A Morello. An addressable quantum dot qubit with fault-tolerant control-fidelity. Nature Nanotechnology, 9(12):981, 2014.

[7] TF Watson, SGJ Philips, Erika Kawakami, DR Ward, Pasquale Scarlino, Menno Veldhorst, DE Savage, MG Lagally, Mark Friesen, and SN Coppersmith. A programmable two-qubit quantum processor in silicon. Nature, 2018.

[8] 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.

[9] Jarryd J Pla, Kuan Y Tan, Juan P Dehollain, Wee H Lim, John JL Morton, Floris A Zwanenburg, David N Jamieson, Andrew S Dzurak, and Andrea Morello. High-fidelity readout and control of a nuclear spin qubit in silicon. Nature, 496(7445):334, 2013.

[10] Charles D Hill, Eldad Peretz, Samuel J Hile, Matthew G House, Martin Fuechsle, Sven Rogge, Michelle Y Simmons, and Lloyd CL Hollenberg. A surface code quantum computer in silicon. Science advances, 1(9):e1500707, 2015.

[11] 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.

[12] John Watrous. THEORY OF QUANTUM INFORMATION. Cambridge University Press, 2018.

[13] Puzzuoli Daniel. Honest approximations to realistic fault models and their applications to efficient simulation of quantum error correction. Master's thesis, University of Waterloo, 2014.

[14] Easwar Magesan. Gaining information about a quantum channel via twirling. Master's thesis, University of Waterloo, 2008.

[15] Man-Duen Choi. Completely positive linear maps on complex matrices. Linear algebra and its applications, 10(3):285–290, 1975.

[16] Daniel Greenbaum. Introduction to quantum gate set tomography. arXiv preprint arXiv:1509.02921, 2015.

[17] Alexei Gilchrist, Nathan K Langford, and Michael A Nielsen. Distance measures to compare real and ideal quantum processes. Physical Review A, 71(6):062310, 2005.

[18] Joel J Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. Physical Review A, 94(5):052325, 2016.

[19] Michael A Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. Physics Letters A, 303(4):249–252, 2002.

[20] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. General teleportation channel, singlet fraction, and quasidistillation. Physical Review A, 60(3):1888, 1999.

[21] Benjamin Schumacher. Sending entanglement through noisy quantum channels. Physical Review A, 54(4):2614, 1996.

[22] George Gabriel Stokes. On the composition and resolution of streams of polarized light from different sources. Transactions of the Cambridge Philosophical Society, 9:399, 1851.

[23] 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.

[24] A Luis and LL Sánchez-Soto. Complete characterization of arbitrary quantum measurement processes. Physical Review Letters, 83(18):3573, 1999.

[25] Alvaro Feito, JS Lundeen, Hendrik Coldenstrodt-Ronge, Jens Eisert, Martin B Plenio, and Ian A Walmsley. Measuring measurement: theory and practice. New Journal of Physics, 11(9):093038, 2009.

[26] 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, 2009.

[27] D Mogilevtsev, J Řeháček, and Z Hradil. Self-calibration for self-consistent tomography. New Journal of Physics, 14(9):095001, 2012.

[28] D Mogilevtsev, J Řeháček, and Z Hradil. Relative tomography of an unknown quantum state. Physical Review A, 79(2):020101, 2009.

[29] D Mogilevtsev, A Ignatenko, Alexander Maloshtan, B Stoklasa, J Rehacek, and Z Hradil. Data pattern tomography: reconstruction with an unknown apparatus. New Journal of Physics, 15(2):025038, 2013.

[30] Merlin Cooper, Michał Karpiński, and Brian J Smith. Local mapping of detector response for reliable quantum state estimation. Nature communications, 5:4332, 2014.

[31] 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.

[32] JF Poyatos, J Ignacio Cirac, and Peter Zoller. Complete characterization of a quantum process: the two-bit quantum gate. Physical Review Letters, 78(2):390, 1997.

[33] Andrew M Childs, Isaac L Chuang, and Debbie W Leung. Realization of quantum process tomography in nmr. Physical Review A, 64(1):012314, 2001.

[34] Jeremy L O'Brien, GJ Pryde, Alexei Gilchrist, DFV James, Nathan K Langford, TC Ralph, and AG White. Quantum process tomography of a controlled-not gate. Physical Review Letters, 93(8):080502, 2004.

[35] M Howard, J Twamley, C Wittmann, T Gaebel, F Jelezko, and J Wrachtrup. Quantum process tomography and linblad estimation of a solid-state qubit. New Journal of Physics, 8(3):33, 2006.

[36] 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.

[37] M Riebe, K Kim, P Schindler, T Monz, PO Schmidt, TK Körber, W Hänsel, H Häffner, CF Roos, and R Blatt. Process tomography of ion trap quantum gates. Physical Review Letters, 97(22):220407, 2006.

[38] Zdenek Hradil. Quantum-state estimation. Physical Review A, 55(3):R1561, 1997.

[39] Jaromír Fiurášek. Maximum-likelihood estimation of quantum measurement. Physical Review A, 64(2):024102, 2001.

[40] Michael A Nielsen, Emanuel Knill, and Raymond Laflamme. Complete quantum teleportation using nuclear magnetic resonance. Nature, 396(6706):52, 1998.

[41] T Yamamoto, M Neeley, E Lucero, RC Bialczak, J Kelly, M Lenander, Matteo Mariantoni, AD O'Connell, D Sank, and H Wang. Quantum process tomography of two-qubit controlled-z and controlled-not gates using superconducting phase qubits. Physical Review B, 82(18):184515, 2010.

[42] 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.

[43] Irene Bongioanni, Linda Sansoni, Fabio Sciarrino, Giuseppe Vallone, and Paolo Mataloni. Experimental quantum process tomography of non-trace-preserving maps. Physical Review A, 82(4):042307, 2010.

[44] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. New Journal of Physics, 12(4):043034, 2010.

[45] Milena D'Angelo, Alessandro Zavatta, Valentina Parigi, and Marco Bellini. Tomographic test of bell's inequality for a time-delocalized single photon. Physical Review A, 74(5):052114, 2006.

[46] Jaroslav Řeháček, Zdeněk Hradil, E Knill, and AI Lvovsky. Diluted maximum-likelihood algorithm for quantum tomography. Physical Review A, 75(4):042108, 2007.

[47] Ferenc Huszár and Neil MT Houlsby. Adaptive bayesian quantum tomography. Physical Review A, 85(5):052120, 2012.

[48] 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.

[49] KS Kravtsov, SS Straupe, IV Radchenko, NMT Houlsby, F Huszár, and SP Kulik. Experimental adaptive bayesian tomography. Physical Review A, 87(6):062122, 2013.

[50] Christopher Granade, Joshua Combes, and DG Cory. Practical bayesian tomography. New Journal of Physics, 18(3):033024, 2016.

[51] 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.

[52] Joseph Emerson, Marcus Silva, Osama Moussa, Colm Ryan, Martin Laforest, Jonathan Baugh, David G Cory, and Raymond Laflamme. Symmetrized characterization of noisy quantum processes. Science, 317(5846):1893–1896, 2007.

[53] Benjamin Lévi, Cecilia C López, Joseph Emerson, and David G Cory. Efficient error characterization in quantum information processing. Physical Review A, 75(2):022314, 2007.

[54] 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.

[55] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. Physical Review Letters, 106(18):180504, 2011.

[56] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. Physical Review A, 85(4):042311, 2012.

[57] Joel J Wallman and Steven T Flammia. Randomized benchmarking with confidence. New Journal of Physics, 16(10):103032, 2014.

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

[59] Jeffrey M Epstein, Andrew W Cross, Easwar Magesan, and Jay M Gambetta. Investigating the limits of randomized benchmarking protocols. Physical Review A, 89(6):062321, 2014.

[60] Joel J Wallman. Randomized benchmarking with gate-dependent noise. arXiv preprint arXiv:1703.09835, 2017.

[61] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout. What randomized benchmarking actually measures. arXiv preprint arXiv:1702.01853, 2017.

[62] Daniel Gottesman. The Heisenberg representation of quantum computers. arXiv preprint quant-ph/9807006, 1998.

[63] M Nest. Classical simulation of quantum computation, the gottesman-knill theorem, and slightly beyond. arXiv preprint arXiv:0811.0898, 2008.

[64] JT Muhonen, A Laucht, S Simmons, JP Dehollain, R Kalra, FE Hudson, S Freer, KM Itoh, DN Jamieson, and JC McCallum. Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. Journal of Physics: Condensed Matter, 27(15):154205, 2015.

[65] Jonas Helsen, Joel J Wallman, Steven T Flammia, and Stephanie Wehner. Multiqubit randomized benchmarking using few samples. arXiv preprint arXiv:1701.04299, 2017.

[66] Zijun Chen, Julian Kelly, Chris Quintana, R Barends, B Campbell, Yu Chen, B Chiaro, A Dunsworth, AG Fowler, and E Lucero. Measuring and suppressing quantum state leakage in a superconducting qubit. Physical Review Letters, 116(2):020501, 2016.

[67] 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.

[68] Antonio D Córcoles, Easwar Magesan, Srikanth J Srinivasan, Andrew W Cross, Matthias Steffen, Jay M Gambetta, and Jerry M Chow. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. Nature communications, 6:6979, 2015.

[69] 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. arXiv preprint arXiv:1310.4492, 2013.

[70] 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(6):062119, 2013.

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

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

[73] Łukasz Rudnicki, Zbigniew Puchała, and Karol Zyczkowski. Gauge invariant information concerning quantum channels. arXiv preprint arXiv:1707.06926, 2017.

[74] Christopher Jackson and Steven van Enk. Nonholonomic tomography. i. the born rule as a connection between experiments. Physical Review A, 95(5):052327, 2017.

[75] Nathaniel Johnston. QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9. `http://qetlab.com`, January 2016.

[76] John Watrous. Simpler semidefinite programs for completely bounded norms. arXiv preprint arXiv:1207.5726, 2012.

[77] Gilles Brassard and Paul Raymond-Robichaud. The equivalence of local-realistic and no-signalling theories. arXiv preprint arXiv:1710.01380, 2017.

[78] Lucien Hardy. Quantum theory from five reasonable axioms. arXiv preprint quant-ph/0101012, 2001.

[79] J Medford, Johannes Beil, JM Taylor, SD Bartlett, AC Doherty, EI Rashba, DP DiVincenzo, H Lu, AC Gossard, and Charles M Marcus. Self-consistent measurement and state tomography of an exchange-only spin qubit. Nature Nanotechnology, 8(9):654, 2013.

[80] Giorgio Brida, Luigi Ciavarella, Ivo Pietro Degiovanni, Marco Genovese, Lapo Lolli, Maria Griselda Mingolla, Fabrizio Piacentini, Mauro Rajteri, Emanuele Taralli, and Matteo GA Paris. Quantum characterization of superconducting photon counters. New Journal of Physics, 14(8):085001, 2012.

[81] R Blume-Kohout, John King Gamble, Erik Nielsen, P Maunz, T Scholten, and K Rudinger. Turbocharging quantum tomography. Technical report, Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States), 2015.

[82] Joel J Wallman. Bounding experimental quantum error rates relative to fault-tolerant thresholds. arXiv preprint arXiv:1511.00727, 2015.

[83] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. Journal of the American statistical association, 58(301):13–30, 1963.

[84] Julian Kelly, R Barends, B Campbell, Y Chen, Z Chen, B Chiaro, A Dunsworth, Austin G Fowler, I-C Hoi, and E Jeffrey. Optimal quantum control using randomized benchmarking. Physical Review Letters, 112(24):240504, 2014.

[85] CJ Ballance, TP Harty, NM Linke, MA Sepiol, and DM Lucas. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. Physical Review Letters, 117(6):060504, 2016.

[86] Salman Beigi and Robert König. Simplified instantaneous non-local quantum computation with applications to position-based cryptography. New Journal of Physics, 13(9):093036, 2011.

APPENDICES

Appendix A

Conversion between representations for quantum maps

Here we derive the formula given in [16] which converts between the Liouville and Choi representations for quantum maps. Note that the formula presented here is slightly different due to the use of a potentially alternative definition of the Choi representation in [16] than the one used here. Recall that the Choi representation is defined for a map Φ in Equation 2.15 as

$$J(\Phi) = \sum_{a,b} \Phi(E_{a,b}) \otimes E_{a,b} \tag{A.1}$$

where $E_{a,b}$ is the standard basis for $2^n \times 2^n$ matrices as before, having entry 1 at the (a, b)-th position and 0 elsewhere. This representation allows for an easy verification of the positivity of a map: recall from Theorem 2.3.1, Φ is CPTP iff $J(\Phi) \geq 0$ and $\mathrm{Tr}_2(J(\Phi)) = \mathbb{1}_{\mathcal{H}^{2^n}}$.

The two formulas are presented as follows: to convert from the Liouville representation to Choi, we use

$$J_\Phi = \sum_{ij} (G_\Phi)_{ij} P_i \otimes P_j^T \tag{A.2}$$

and backwards,

$$G_\Phi = \sum_{i,j} \mathrm{Tr}\{J_\Phi (P_i \otimes P_j^T)\} E_i \otimes E_j \tag{A.3}$$

with a slight abuse of notation, where we re-labeled the standard basis with a single subscript which is ordered by the row and then column of the old labels. For example, for 2×2 real matrices, we re-label E_{11} as E_1 and E_{21} as E_3, etc. We will denote the set of these re-labeled standard bases for n-qubits as $\mathbf{E}_n$.

Here we give a derivation of Equation A.2, and A.3 should follow directly by the orthonormality of Pauli basis matrices. To show the desired formula, it is useful to first express the resultant state from acting Φ on an arbitrary operator A in terms of its Liouville coefficients. This can be easily done using Equation 2.22: in particular,

$$\Phi(A) = \sum_i (G_\Phi |\rho\rangle\rangle)_i P_i = \sum_{ij} P_i \operatorname{Tr}[P_i \Phi(P_j)] \operatorname{Tr}[A P_j] \tag{A.4}$$

Next, we will need the conversion between the Pauli basis and the standard basis: by collecting the Pauli and the standard bases in a vector form, one can succinctly write down the conversion in the following matrix form

$$\vec{P} := U\vec{E},\ \vec{E} = U^\dagger \vec{P} \tag{A.5}$$

where the components of $\vec{P}$ and $\vec{E}$ are elements of $\mathbf{P}_n$ and $\mathbf{E}_n$, and U is a unitary matrix because both $\vec{P}$ and $\vec{E}$ are orthonormal basis for the vector space $L(\mathcal{H}^{2^n})$, and the transformation matrix between two orthonormal bases is unitary. In particular, the rows and columns of U are pairwise linearly-independent, so the following relationship holds:

$$\sum_j U^*_{ij} U_{kj} = \sum_j U^*_{ji} U_{jk} = \delta_{ik} \tag{A.6}$$

The above also implies that

$$E_i = \sum_j (U^\dagger_{ij}) P_j = \sum_j U^*_{ji} P_j \tag{A.7}$$

With these we will express E_i and $\Phi(E_i)$ in a desired form. First, since all E's are real, we have

$$E_i = \bar{E}_i = \sum_j \overline{U^*_{ji} P_j} = \sum_j U_{ji} P_j^T \tag{A.8}$$

where the last line is due to $\bar{P}_j = P_j^T\ \forall j$.

Second, using Equation A.4 and A.7, we have

$$\begin{aligned}\Phi(E_i) &= \sum_{jkl} P_j \operatorname{Tr}[P_j \Phi(P_k)] \operatorname{Tr}[U_{li}^* P_l P_k] \\ &= \sum_{jkl} U_{li}^* \operatorname{Tr}[P_j \Phi(P_k)] \delta_{lk} P_j \\ &= \sum_{jk} U_{ki}^* \operatorname{Tr}[P_j \Phi(P_k)] P_j\end{aligned} \tag{A.9}$$

Plug the two above equation into the definition for J_Φ, we have

$$\begin{aligned}J_\Phi &= \sum_i \Phi(E_i) \otimes E_i \\ &= \sum_{ijkl} U_{ki}^* U_{li} \operatorname{Tr}[P_j \Phi(P_k)] P_j \otimes P_l^T \\ &= \sum_{jkl} \delta_{kl} \operatorname{Tr}[P_j \Phi(P_k)] P_j \otimes P_l^T \\ &= \sum_{jk} \operatorname{Tr}[P_j \Phi(P_k)] P_j \otimes P_k^T \\ &= \sum_{jk} (G_\Phi)_{jk} P_j \otimes P_k^T\end{aligned} \tag{A.10}$$

which is the desired formula, Equation A.2.