Honest Approximations to Realistic Fault Models and Their Applications to Efficient Simulation of Quantum Error Correction

by

Daniel Puzzuoli

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Abstract

Understanding the performance of realistic noisy encoded circuits is an important task for the development of large-scale practical quantum computers. Specifically, the development of proposals for quantum computation must be well informed by both the qualities of the low-level physical system of choice, and the properties of the high-level quantum error correction and fault-tolerance schemes. Gaining insight into how a particular computation will play out on a physical system is in general a difficult problem, as the classical simulation of arbitrary noisy quantum circuits is inefficient. Nevertheless, important classes of noisy circuits *can* be simulated efficiently. Such simulations have led to numerical estimates of threshold errors rates and resource estimates in topological codes subject to efficiently simulable error models.

This thesis describes and analyzes a method that my collaborators and I have introduced for leveraging efficient simulation techniques to understand the performance of large quantum processors that are subject to errors lying *outside* of the efficient simulation algorithm's applicability. The idea is to approximate an arbitrary gate error with an error from the efficiently simulable set in a way that "honestly" represents the original error's ability to preserve or distort quantum information. After introducing and analyzing the individual gate approximation method, its utility as a means for estimating circuit performance is studied. In particular, the method is tested within the use-case for which it was originally conceived; understanding the performance of a hypothetical physical implementation of a quantum error-correction protocol. It is found that the method performs exactly as desired in all cases. That is, the circuits composed of the approximated error models honestly represent the circuits composed of the errors derived from the physical models.

Acknowledgements

First and foremost I would like to thank David Cory and Joseph Emerson for giving me a chance to work in this field. Without them I don't know where I'd be; probably on some street corner selling pencils from a cup.

Next I'd like to thank Easwar Magesan, both for his collaboration, and for the guidance and support that only someone far enough – but not *too* far – ahead of you can provide. My office mate and collaborator, Christopher Granade, has been indispensable in our research, as well as in my continual re-education of things oft-forgotten, and sometimes never-remembered. My other collaborators, Holger Haas and Ben Criger, have enriched my graduate school life, not least of all through things which have nothing to do with our work.

Lastly I'd like to thank all of the members of David Cory's research lab, who have livened the experience. Honourable mention goes to Ian Hincks, who has been a good friend and excellent source of knowledge regarding Middle Earth.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Quantum computers offer potentially exponential speed-ups over their classical counterparts for well known and important computational problems [2, 3, 4]. As such, one of the biggest tasks of applied quantum information is building a functioning quantum computer. Doing so requires input at all levels; a deep understanding of the low-level physical system, of high-level error correction protocols, and of how to effectively connect the two.

Connecting the low and high levels necessarily requires analysis of how the physical system and error-correction protocol will work together. With regards to circuit-model based computations, such an analysis can be phrased as the question: how will a given circuit perform on a given physical system? This of course, is a very hard problem to approach, as one cannot simply simulate the dynamics of the physical system on an existing (classical) computer efficiently.

Efforts to analytically understand the performance of noisy quantum circuits have resulted in threshold theorems for fault-tolerant circuits (e.g., [5]). That is, for certain fault-tolerant error-correction protocols, given certain assumptions on the noise model, it is possible to analytically find a threshold "error strength" for which, if the "error strength" is below this threshold, the fault tolerance scheme can simulate a noiseless circuit with arbitrary precision (at the cost of additional physical resources)[1].

For some error-correction schemes, no analytic theorem currently exists. Nevertheless, one would still like to get an idea of thresholds; how small do the errors have to be for a computation to succeed? In some cases, efficiently simulable *sub-theories* of quantum mechanics have proven useful for investigating such questions. If one restricts the types of

[1]The "error strength" is generally some parameter or set of parameters that quantifies the error.

operations, initial states, and error models in an intelligent way, simulating the performance of a quantum circuit can become tractable (see Section 3.2). On first glance one might expect that the types of circuits admitting efficient simulation will be extremely limited, but it turns out that many important and interesting circuits can be implemented using only efficiently simulable circuit elements. The most well-known example is stabilizer circuits, which are efficiently simulable on stabilizer states by the Gottesman-Knill Theorem (see Section 3.2). Stabilizer circuits include error-correction steps in stabilizer codes, and can be promoted to universal quantum computation by allowing the preparation of particular non-stabilizer states [6]. Thus, while stabilizer circuits (on stabilizer states) are not universal, and therefore will not be the only operations used in a meaningful computation, it is generally believed that the majority of operations implemented on a quantum computer will be devoted to error-correction, and so understanding the performance of circuits containing only these elements remains valuable.

Such efficient simulation techniques have provided numerical estimates to thresholds for topological codes [7, 8]. Even in cases when threshold theorems are known, such as in the surface code [9], simulation can be used to find more optimistic estimates of a threshold value [1]. They have also been useful for understanding resource requirements in cases when the noise strength is already well below the threshold [10]. Thus, given these applications, it would be highly desirable to find some way of using efficient simulation techniques to say something about the performance of circuits on particular physical machines. The difficulty of course arrises from the fact that the errors on gates implemented on a real system will in general not take an efficiently simulable form.

1.1 Honest Representations of Channels

The idea presented in this thesis, which was proposed in [11], is a method for applying classical efficient simulation techniques to noise models that lie outside of the efficient simulation formalism. Any imperfectly implemented circuit can be viewed as a series of perfect gates intermixed with noise operations. Concretely, we can write an imperfect implementation Λ (a completely positive and trace preserving - CPTP - map) of a unitary U, as $\Lambda = \mathcal{E} \circ U$, where $\mathcal{E}$ is some CPTP map that represents the noise of Λ as some operation occurring after U. If, as discussed earlier, we are interested in circuits which are efficiently simulable according to some algorithm, U will naturally be chosen from the efficiently simulable set. However, the error $\mathcal{E}$ will almost certainly not be efficiently simulable, and thus Λ will not be.

From here, our idea is simple. Denoting the efficiently simulable set of channels as S, we

want to find some $\Gamma \in S$ that approximates the error $\mathcal{E}$. As we're interested in evaluating circuit performance however, it isn't enough that Γ is *close* to $\mathcal{E}$ according to some metric. We also want Γ to be an *honest* representation of the error $\mathcal{E}$, in that we never want Γ to underestimate the strength of the error $\mathcal{E}$ on the state. We want this property to be *global*, in the sense that we want Γ to always be somehow worse than $\mathcal{E}$ regardless of the state that is sent through.

The hope then is that if we can honestly represent every error in the circuit, then the circuit composed of the honest representations will also be an honest representation of the original circuit. We refer to this as the problem of "error composition"[2].

We quantify error using the distance induced by the Schatten p-norms $\|\cdot\|_p$ (defined in Chapter 2) on quantum states. We are mainly interested in the case when $p = 1$, as there is a strong operational interpretation of the quantity $\|\rho_0 - \rho_1\|_1$ in terms of the *distinguishability* of two quantum states ρ_0 and ρ_1, discussed in detail in Chapter 3. Thus, if we feed a state ρ through an error $\mathcal{E}$, we quantify the error by $\|\rho - \mathcal{E}(\rho)\|_p$, for some $p \geq 1$, where, again, we are most interested in the case when $p = 1$, which measures how hard it is to tell whether or not the error has occurred. We say that another error Γ is an *honest representation* of $\mathcal{E}$ if, for some $p \geq 1$, $\|\rho - \Gamma(\rho)\|_p \geq \|\rho - \mathcal{E}(\rho)\|_p$ for all pure states ρ. That is, no matter what pure state ρ is used, Γ always moves ρ further than $\mathcal{E}$. Note that we use the term "honest" differently from [11, 12], where "honesty" is reserved solely for the $p = 1$ case. We remove this restriction here for simplicity, to avoid specifying which p is meant when it is clear from context. While the cases when $p > 1$ have no operational interpretation (known to me), if this type of inequality is satisfied for all pure states ρ, then clearly Γ is somehow *globally* worse than $\mathcal{E}$.

1.2 Overview

This thesis is organized as follows:

- Chapter 2 - Notation and relevant facts that will be used in this thesis are established.
- Chapter 3 - The concept of distinguishability is reviewed, and efficiently simulable sub-theories of quantum mechanics are discussed.

[2]Note that we use the word "composition" to refer to both direct (serial) composition of maps, as well as tensor products (parallel composition) of maps.

- Chapter 4 - Results that are useful for finding approximations to arbitrary CPTP maps that are also honest are established. Specifically, the statement of honesty given in this introduction has the "for all pure states" qualification. Sufficient conditions on the form of the maps which are easy to compute, and can therefore be used in optimization problems for finding honest approximations, are given. An approximation problem based on these conditions is then specified, and then applied to some instructive qubit examples.

- Chapter 5 - The utility of honest approximations for judging circuit performance is numerically investigated. We begin with continuous-time models of physical systems, build gates from them, and then test our method on a typical error correcting circuit using these gates. We find somewhat remarkably that in all cases, our approximations perform exactly as desired. That is, the circuits composed of the honest approximations are themselves honest representations of the circuits composed with the errors derived from the physical models. This evidence is a strong boon for our method. This chapter is a reproduction of the content of [12].

- Chapter 6 - The performance of "Pauli twirling" is analyzed (see Section 2.3), which is a method of mapping an arbitrary error to a Pauli channel (which is efficiently simulable via the formalism we focus on in this thesis, see Chapter 3). Pauli twirling has been investigated as another means for performing the same task that our method is designed for, due to that fact that it has certain computational benefits. In Chapter 5 we include statistics on the performance of Pauli twirling, which are analyzed in this chapter.

- Chapter 7 - Concluding statements are given, remaining open questions are discussed, and possible future directions and next steps are considered.

Chapter 2

Mathematical Background

First, we establish some basic notation and concepts to work from. This notation is primarily borrowed from the course notes of Watrous [13].

2.1 Basic Objects and Notation

2.1.1 Complex Inner Product Space

As we are dealing with finite-dimensional quantum systems, we wish to work in complex finite-dimensional Hilbert spaces $\mathcal{X}$ that are isomorphic to $\mathbb{C}^n$ (with $n = \dim(\mathcal{H})$) with the standard Euclidean inner product $\langle \cdot, \cdot \rangle$, and as such, we will simply work in $\mathbb{C}^n$. A vector $v \in \mathbb{C}^n$ is an n-tuple of complex numbers (a column vector), with brackets and subscripts denoting entries. (Eg. $(v)_i$ is the i^{th} entry of $v \in \mathbb{C}^n$. This notation is chosen to avoid notational collisions when a collection of vectors is indexed with subscripts.) The inner product of two elements $x, y \in \mathbb{C}^n$ is given by $\langle x, y \rangle = \sum_{i=1}^{d} (x)_i^* (y)_i$, where * denotes complex conjugation. Using standard definitions of matrix multiplication, transposition (^{T}), and conjugate transposition (†), we can write $\langle x, y \rangle = x^\dagger y$ (if $x, y \in \mathbb{R}^n$, then $\langle x, y \rangle = x^T y$). For a given vector space $\mathbb{C}^n$, we call the set of vectors $\{e_i\}_{i=1}^n$ satisfying $(e_i)_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta function, the *standard basis* for $\mathbb{C}^n$.

To represent a vector $x \in \mathbb{C}^n$, we will sometimes use Dirac notation, $|x\rangle$. In Dirac notation, we use the conventions $|x\rangle^\dagger \equiv \langle x|$, $|x\rangle^T \equiv \langle x^*|$, and $\langle x|y\rangle \equiv \langle x, y\rangle$. When there is no need to make the dimension of the space explicit, we will just use symbols like $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ to refer to finite-dimensional Hilbert spaces.

2.1.2 Linear Operators and Hilbert-Schmidt Inner Product

The set of linear operators mapping $\mathbb{C}^n$ to $\mathbb{C}^m$ is denoted $\mathrm{L}(\mathbb{C}^n, \mathbb{C}^m)$. We use the simplifying notation $\mathrm{L}(\mathbb{C}^n) = \mathrm{L}(\mathbb{C}^n, \mathbb{C}^n)$. These operators can be represented by matrices, where we denote the i, j^{th} entry of a matrix A as $(A)_{ij} = \langle f_i, Ae_j \rangle$, where $f_i \in \mathbb{C}^m$ and $e_j \in \mathbb{C}^n$ are elements of the standard bases. The Hilbert-Schmidt inner product on operators $A, B \in \mathrm{L}(\mathbb{C}^n, \mathbb{C}^m)$ is defined as $\langle A, B \rangle = \mathrm{Tr}(A^\dagger B)$, where Tr is the standard trace operation. We will use the symbol $\mathbb{1}$ to refer to the identity element in $\mathrm{L}(\mathbb{C}^n)$. If there is any chance of confusion, it will be labelled as $\mathbb{1}_n$, or if the dimension of the space is not specified (ie. we are using a symbol like $\mathcal{X}$ to refer to the space), we will label it as $\mathbb{1}_{\mathcal{X}}$.

Of primary importance for us are Hermitian operators, which satisfy $A = A^\dagger$, or in index notation, $(A)_{ij} = (A)^*_{ji}$. We denote the set of Hermitian operators in $\mathrm{L}(\mathbb{C}^n)$ as $\mathrm{Herm}(\mathbb{C}^n)$. A nice property for $A, B \in \mathrm{Herm}(\mathbb{C}^n)$ which can be easily verified is $\langle A, B \rangle \in \mathbb{R}$. Another easy to verify property is that $\mathrm{Herm}(\mathbb{C}^n)$ is closed under addition and multiplication by real scalars. Thus, when $\mathrm{Herm}(\mathbb{C}^n)$ is thought of as a vector space over $\mathbb{R}$, it is a real inner product space when coupled with the Hilbert-Schmidt inner product. As it takes n^2 real numbers to specify an element in $\mathrm{Herm}(\mathbb{C}^n)$, it is an n^2 dimensional real inner product space. It can be shown that $A \in \mathrm{Herm}(\mathbb{C}^n)$ if and only if $\langle x, Ax \rangle \in \mathbb{R}$ for all $x \in \mathbb{C}^n$.

Also of importance is the set of positive semi-definite operators. An operator $P \in \mathrm{L}(\mathbb{C}^n)$ is called positive semi-definite if $\langle x, Px \rangle \geq 0$ for all $x \in \mathbb{C}^n$. We denote the set of such operators as $\mathrm{Pos}(\mathbb{C}^n)$, and use the notation $P \geq 0$ as equivalent to specifying that $P \in \mathrm{Pos}(\mathbb{C}^n)$. It follows from $\langle x, Px \rangle \in \mathbb{R}$ that $\mathrm{Pos}(\mathcal{X}) \subset \mathrm{Herm}(\mathcal{X})$. For two operators $P, Q \in \mathrm{Herm}(\mathbb{C}^n)$, we write $P \geq Q$ if $P - Q \geq 0$. A nice property for $P, Q \in \mathrm{Pos}(\mathbb{C}^n)$, is that $\langle P, Q \rangle \geq 0$. An important subset of $\mathrm{Pos}(\mathbb{C}^n)$ is the set of *density* operators, which represent quantum states. We denote the set of density operators as

$$\mathrm{D}(\mathbb{C}^n) \equiv \{\rho \in \mathrm{Pos}(\mathbb{C}^n) : \mathrm{Tr}(\rho) = 1\}.$$

A final subset of $\mathrm{Herm}(\mathbb{C}^n)$ that we give a special symbol to is

$$\mathrm{Herm}_1(\mathbb{C}^n) = \{A \in \mathrm{Herm}(\mathbb{C}^n) : \mathrm{Tr}(A) = 1\}.$$

2.1.3 Super Operators and Quantum Channels

A "super operator" is a linear operator that maps $\mathrm{L}(\mathbb{C}^n)$ to $\mathrm{L}(\mathbb{C}^m)$. We denote the set of such operators as $\mathrm{T}(\mathbb{C}^n, \mathbb{C}^m)$, and as before we write $\mathrm{T}(\mathbb{C}^n) = \mathrm{T}(\mathbb{C}^n, \mathbb{C}^n)$. Similarly to how we use $\mathbb{1}$ to refer to the identity linear operator, with subscripts denoting which space it acts on, we will use $\mathbb{1}_{\mathrm{L}(\mathbb{C}^n)}$ to refer to the identity super operator in $\mathrm{T}(\mathbb{C}^n)$.

For a map $\Lambda \in \mathrm{T}(\mathbb{C}^n, \mathbb{C}^m)$, we have the following classifications:

- Λ is *trace preserving* (TP) if for all $A \in \mathrm{L}(\mathbb{C}^n)$, $\mathrm{Tr}(\Lambda(A)) = \mathrm{Tr}(A)$.
- Λ is *Hermiticity preserving* if for $A \in \mathrm{Herm}(\mathbb{C}^n)$, $\Lambda(A) \in \mathrm{Herm}(\mathbb{C}^m)$.
- Λ is *positive* if for $A \in \mathrm{Pos}(\mathbb{C}^n)$, $\Lambda(A) \in \mathrm{Pos}(\mathbb{C}^m)$. Note that positive maps are necessarily also Hermiticity preserving by linearity and the fact that any Hermitian matrix can be written as the difference of two positive matrices.
- Λ is *completely positive* (CP) if, for any $k \in \mathbb{N}$, $\Lambda \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^k)} \in \mathrm{T}(\mathbb{C}^n \otimes \mathbb{C}^k, \mathbb{C}^m \otimes \mathbb{C}^k)$ is positive, where $\mathbb{1}_{\mathrm{L}(\mathbb{C}^k)} \in \mathrm{T}(\mathbb{C}^k)$ is the identity super operator.
- Λ is called a *quantum channel* if it is both CP and TP (CPTP). We denote the set of CPTP maps from $\mathrm{L}(\mathbb{C}^n)$ to $\mathrm{L}(\mathbb{C}^m)$ as $\mathrm{C}(\mathbb{C}^n, \mathbb{C}^m)$ and denote $\mathrm{C}(\mathbb{C}^n) = \mathrm{C}(\mathbb{C}^n, \mathbb{C}^n)$.

A quantum channel that we use often is the partial trace, $\mathrm{Tr}_{\mathcal{Y}} \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{X})$, defined as $\mathrm{Tr}_{\mathcal{Y}}(A \otimes B) = \mathrm{Tr}(B)A$, for all $A \in \mathrm{L}(\mathcal{X})$, $B \in \mathrm{L}(\mathcal{Y})$, and extended to all operators in $\mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ by linearity. The subscript denotes which subsystem the trace is being applied to.

2.1.4 Characterizations of CPTP maps

In this document we make use of two characterizations of CPTP maps, the Kraus (or operator-sum) and Stinespring representations. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite dimensional complex inner product spaces. For $\Lambda \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the following statements are equivalent:

1. $\Lambda \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, ie. Λ is a CPTP map.
2. There exists operators $\{A_i\}_{i=1}^k \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$, satisfying $\sum_{i=1}^k A_i^\dagger A_i = \mathbb{1}_{\mathcal{X}}$ for which $\Lambda(X) = \sum_{i=1}^k A_i X A_i^\dagger$ for all $X \in \mathrm{L}(\mathcal{X})$. The set $\{A_i\}_{i=1}^k$ is called a *Kraus representation* for Λ.
3. There exists another finite dimensional complex inner product space $\mathcal{Z}$, and operator $B \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for which $\Lambda(X) = \mathrm{Tr}_{\mathcal{Z}}(BXB^\dagger)$ for all $X \in \mathrm{L}(\mathcal{X})$, and $B^\dagger B = \mathbb{1}_{\mathcal{X}}$. We call the tuple $(\mathcal{Z}, B)$ a *Stinespring representation* for Λ.

2.1.5 Schatten p-norms

In this document we will make use of Schatten p-norms on elements of $\mathrm{L}(\mathcal{X}, \mathcal{Y})$. For an element $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $1 \leq p < \infty$, $\|\cdot\|_p$ is defined as

$$\|A\|_p = (\sum_i \sigma_i^p)^{\frac{1}{p}}, \tag{2.1}$$

where $\{\sigma_i\}$ are the singular values of A (see Appendix A). As we will use these norms exclusively on Hermitian operators, we will simply write, for $A \in \mathrm{Herm}(\mathcal{X})$,

$$\|A\|_p = (\sum_i |\lambda_i|^p)^{\frac{1}{p}}, \tag{2.2}$$

where $\{\lambda_i\}$ are the eigenvalues of A. Of primary importance for us are the special cases $p = 1$ and $p = 2$, which, for $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, take the form

$$\|A\|_1 = \mathrm{Tr}(\sqrt{A^\dagger A}) \tag{2.3}$$

$$\|A\|_2 = \sqrt{\mathrm{Tr}(A^\dagger A)}. \tag{2.4}$$

An important property of $\|\cdot\|_1$ that we make use of is the monotonicity of $\|\cdot\|_1$ under the partial trace. That is, for $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$, it holds that

$$\|X\|_1 \geq \|\mathrm{Tr}_{\mathcal{Y}}(X)\|_1 . \tag{2.5}$$

Another useful relation is that the one-norm distance between unit vectors $u, v \in \mathcal{X}$ is given simply as $\|uu^\dagger - vv^\dagger\|_1 = 2\sqrt{1 - |\langle u, v\rangle|^2}$.

2.1.6 The Fidelity Function and Purification

We make use of the two related notions of fidelity and purification. For two positive operators $A, B \in \mathrm{Pos}(\mathcal{X})$, the fidelity is defined as the function $\mathrm{F}(A, B) = \|\sqrt{A}\sqrt{B}\|_1$. If one of the operators is a rank 1 projector, ie. $A = uu^\dagger$ for some unit vector $u \in \mathcal{X}$, then this definition reduces to $\mathrm{F}(uu^\dagger, B) = \sqrt{|\langle u, Bu\rangle|}$, and similarly, for $B = vv^\dagger$, $\mathrm{F}(uu^\dagger, vv^\dagger) = |\langle u, v\rangle|$. Given the simple forms of both the one-norm distance and the fidelity for pairs of pure states, one can see that for unit vectors $u, v \in \mathcal{X}$,

$$\|uu^\dagger - vv^\dagger\|_1 = 2\sqrt{1 - F(uu^\dagger, vv^\dagger)^2}. \tag{2.6}$$

For a state $\rho \in \mathrm{D}(\mathcal{X})$, a *purification* is any pure state in some composite system $u \in \mathcal{X} \otimes \mathcal{Y}$ for which $\rho = \mathrm{Tr}_{\mathcal{Y}}(uu^\dagger)$. A state $\rho \in \mathrm{D}(\mathcal{X})$ can be purified to a vector in $\mathcal{X} \otimes \mathcal{Y}$ if and only if $\dim(\mathcal{Y}) \geq \mathrm{rank}(\rho)$. A useful theorem called Uhlmann's Theorem [14] connects the concepts of fidelity and purification.

Theorem 2.1.1. *Given two states $\rho_1, \rho_2 \in L(\mathcal{X})$, it holds that for $dim(\mathcal{Y}) = \max\{rank(\rho_1), rank(\rho_2)\}$, and any purification $u \in \mathcal{X} \otimes \mathcal{Y}$ of ρ_1*

$$F(\rho_1, \rho_2) = \max\{|\langle u, v\rangle| : v \in \mathcal{X} \otimes \mathcal{Y} \text{ is a purification of } \rho_2\} \tag{2.7}$$

A clear consequence of this theorem is that for $\rho_1, \rho_2 \in \mathrm{D}(\mathcal{X})$ with purifications $u, v \in \mathcal{X} \otimes \mathcal{Y}$, $\mathrm{F}(\rho_1, \rho_2) \geq \mathrm{F}(uu^\dagger, vv^\dagger)$. Combining Uhlmann's theorem with the monotonicity of the 1-norm under partial trace, $\|\rho_1 - \rho_2\|_1 \leq \|uu^\dagger - vv^\dagger\|_1$, one is able to arrive at one of the Fuchs-van de Graaf inequalities [15]:

$$\|\rho_1 - \rho_2\|_1 \leq 2\sqrt{1 - \mathrm{F}(\rho_1, \rho_2)^2}. \tag{2.8}$$

2.2 Vectorization and the Generalized Bloch Representation

For any finite dimensional complex vector space $\mathcal{X} = \mathbb{C}^n$, operators in $\mathrm{L}(\mathcal{X})$ also form a complex finite dimensional vector space (with $\dim(\mathrm{L}(\mathcal{X})) = \dim(\mathcal{X})^2$). Thus, we can represent operators in $\mathrm{L}(\mathcal{X})$ themselves as column vectors, and linear maps in $\mathrm{T}(\mathcal{X}, \mathcal{Y})$, for some complex vector space $\mathcal{Y} = \mathbb{C}^m$, as matrices. Explicitly, this can be done by choosing orthonormal bases $B_{\mathcal{X}} = \{\lambda_i\}_{i=1}^{n^2}$ and $B_{\mathcal{Y}} = \{\sigma_i\}_{i=1}^{m^2}$ for $\mathrm{L}(\mathcal{X})$ and $\mathrm{L}(\mathcal{Y})$ respectively, and defining $\mathrm{vec} : \mathrm{L}(\mathcal{X}) \to \mathbb{C}^{n^2}$ entry wise as

$$(\mathrm{vec}(A))_i = \langle \lambda_i, A\rangle,$$

and similarly for $\mathrm{vec} : \mathrm{L}(\mathcal{Y}) \to \mathbb{C}^{m^2}$. (We use "vec" to refer to all such vectorization functions. Context and what operators it is being applied to provide enough information on which particular "vec" function is meant.) By straightforward computation, it can be seen that vec is an isometry. That is, for $A, B \in \mathrm{L}(\mathcal{X})$, $\langle \mathrm{vec}(A), \mathrm{vec}(B)\rangle = \langle A, B\rangle$.

In this way, the vectorized, or matrix, representation of a map $\Lambda \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, which we denote $S_\Lambda \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$, can be defined as the operator for which

$$S_\Lambda \mathrm{vec}(X) = \mathrm{vec}(\Lambda(X)),$$

for all $X \in \mathrm{L}(\mathcal{X})$. Entry wise, it is defined as

$$(S_\Lambda)_{ij} = \langle \mathrm{vec}(\sigma_i), S_\Lambda \mathrm{vec}(\lambda_j)\rangle = \langle \mathrm{vec}(\sigma_i), \mathrm{vec}(\Lambda(\lambda_j))\rangle = \langle \sigma_i, \Lambda(\lambda_j)\rangle.$$

2.2.1 Bloch Bases

We define a *Bloch Basis* for $\mathrm{L}(\mathcal{X} = \mathbb{C}^n)$, $\{\lambda_i\}_{i=0}^{n^2-1}$, as an orthonormal basis for $\mathrm{L}(\mathcal{X})$ (with the Hilbert-Schmidt inner product) that satisfies the following additional properties.

- $\lambda_0 = \frac{1}{\sqrt{n}}\mathbb{1}_{\mathcal{X}}$.
- $\{\lambda_i\}_{i=0}^{n^2-1} \subset \mathrm{Herm}(\mathcal{X})$.

In other words, a Bloch basis for $\mathrm{L}(\mathcal{X})$ is an orthonormal basis consisting of Hermitian operators that are all traceless except for the element labeled with 0, which is a normalized identity operator.

For $\mathcal{X} = \mathbb{C}^2$, an example of such a basis is the normalized Pauli operators, given by:

$$\lambda_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \lambda_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \lambda_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \lambda_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Similarly, for $\mathcal{X} = \mathbb{C}^{2^n}$, one can take the basis consisting of all normalized n-fold tensor products of Pauli operators. For arbitrary dimension, an explicit construction of a Bloch basis is given in Appendix A.2.

2.2.2 Properties of Matrix Representations of Channels in Bloch Bases

As noted earlier, for $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} = \mathbb{C}^m$, and respective Bloch bases $\{\lambda_i\}_{i=0}^{n^2-1}$ and $\{\sigma_i\}_{i=0}^{m^2-1}$, the matrix representation of a map $\Lambda \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, S_Λ is defined entry wise as $(S_\Lambda)_{ij} = \langle \sigma_i, \Lambda(\lambda_i)\rangle$.

Now, if we assume that $\Lambda \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, then it is Hermiticity preserving, and it follows that $(S_\Lambda)_{ij} \in \mathbb{R}$ for all i, j, and so $S_\Lambda \in \mathbb{R}^{m^2 \times n^2}$ (a real matrix with m^2 rows and n^2 columns). Λ will also be trace preserving, and as such

$$(S)_{0j} = \langle \sigma_0, \Lambda(\lambda_j)\rangle = \frac{1}{\sqrt{m}}\mathrm{Tr}(\Lambda(\lambda_j)) = \frac{1}{\sqrt{m}}\mathrm{Tr}(\lambda_j) = \sqrt{\frac{n}{m}}\delta_{0j}.$$

Thus, we can write the matrix representation S_Λ of any $\Lambda \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ as

$$S_\Lambda = \left(\begin{array}{c|c} \sqrt{n/m} & 0^T \\ \hline v & R \end{array}\right),$$

where 0 is the zero vector in $\mathbb{R}^{n^2-1}$, $v \in \mathbb{R}^{m^2-1}$ and $R \in \mathbb{R}^{m^2-1 \times n^2-1}$. (Here we use a block matrix notation, where to the left of the vertical line is a single column and to the right is n^2-1 columns, and above the horizontal line is one row and below is m^2-1 rows.)

2.2.3 The Generalized Bloch Representation of Quantum States and Quantum Channels

Using the details of the previous sections, we now define the generalized Bloch representation of quantum states and channels. The aim of this representation is to represent states $\rho \in \mathrm{D}(\mathcal{X} = \mathbb{C}^n)$ as vectors in $\mathrm{S}(\mathbb{R}^{n^2-1}) = \{r \in \mathbb{R}^{n^2-1} | \|r\|_2 \leq 1\}$, and channels $\Lambda \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ as linear transformations from $\mathbb{R}^{n^2-1}$ to $\mathbb{R}^{m^2-1}$ that take $\mathrm{S}(\mathbb{R}^{n^2-1})$ into $\mathrm{S}(\mathbb{R}^{m^2-1})$.

Let $B_{\mathcal{X}} = \{\lambda_i\}_{i=0}^{n^2-1}$ be a Bloch basis for $\mathrm{L}(\mathcal{X})$. Define the function $\mathrm{bvec} : \mathrm{Herm}_1(\mathcal{X}) \to \mathbb{R}^{n^2-1}$ as

$$\mathrm{bvec}(X) = \frac{1}{\alpha(n)}(\langle \lambda_1, X \rangle, ..., \langle \lambda_{n^2-1}, X \rangle)^T$$

for all $X \in \mathrm{Herm}_1(\mathcal{X})$, where $\alpha(n)$ is a to-be-determined positive function of the dimension n. For $X \in \mathrm{Herm}_1(\mathbb{C}^n)$, we call $\mathrm{bvec}(X)$ the *(generalized) Bloch vector* of X. Note that for $X \in \mathrm{Herm}_1(\mathcal{X})$, the bvec and vec functions are related as $\mathrm{vec}(X) = \left(\sqrt{1/n} \;\middle|\; \alpha(n)\mathrm{bvec}(X)^T \right)^T$.

We now determine a form for $\alpha(n)$ that will fulfill the desire that for $\rho \in \mathrm{D}(\mathcal{X})$, $r = \mathrm{bvec}(\rho) \in \mathrm{S}(\mathbb{R}^{n^2-1})$. We do this by choosing $\alpha(n)$ so that $\|r\|_2 = 1$ if ρ is pure. If ρ is pure, then $\mathrm{Tr}(\rho^2) = \mathrm{Tr}(\rho) = 1$. Thus

$$1 = \mathrm{Tr}(\rho^2) = \langle \rho, \rho \rangle = \langle \mathrm{vec}(\rho), \mathrm{vec}(\rho) \rangle = \frac{1}{n} + \alpha(n)^2 \|r\|_2^2. \tag{2.9}$$

Setting $\|r\|_2 = 1$ and solving for $\alpha(n)$, we find that $\alpha(n) = \sqrt{\frac{n-1}{n}}$. Thus, for pure states, $\|\mathrm{bvec}(\rho)\| = 1$, and it follows by linearity of vec and bvec, the triangle inequality, and that all density operators are convex combinations of pure states, that $\|\mathrm{bvec}(\rho)\|_2 \leq 1$ in general. Thus, for $X \in \mathrm{Herm}_1(\mathcal{X})$, and $r = \mathrm{bvec}(X)$, we have that

$$X = \frac{\mathbb{1}_{\mathcal{X}} + \sqrt{n(n-1)} r \cdot \lambda}{n} \tag{2.10}$$

and

$$\mathrm{vec}(X) = \begin{pmatrix} \sqrt{1/n} \\ \sqrt{\frac{(n-1)}{n}} r \end{pmatrix}, \tag{2.11}$$

where $r \cdot \lambda = \sum_{i=1}^{n^2-1} (r)_i \lambda_i$.

The *(generalized) Bloch representation* of a channel $\Lambda \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is a tuple (M, t), where $M \in \mathbb{R}^{(m^2-1)\times(n^2-1)}$ and $t \in \mathbb{R}^{m^2-1}$, defined in the following way. For all $X \in \mathrm{Herm}_1(\mathcal{X})$,

$$\mathrm{bvec}(\Lambda(X)) = M\mathrm{bvec}(X) + t. \tag{2.12}$$

Recall that the matrix representation of a channel Λ, denoted S_Λ, is given by

$$S = \left(\begin{array}{c|c} \sqrt{n/m} & 0^T \\ \hline v & R \end{array}\right), \tag{2.13}$$

for some $v \in \mathbb{R}^{m^2-1}$ and $R \in \mathbb{R}^{(m^2-1)\times(n^2-1)}$. We claim and then verify that the following definitions of M and t satisfy Equation (2.12):

$$M = (\sqrt{\frac{m-1}{m}}/\sqrt{\frac{n-1}{n}})^{-1} R \tag{2.14}$$

$$\vec{t} = (\sqrt{n}\sqrt{\frac{m-1}{m}})^{-1} \vec{v}. \tag{2.15}$$

Writing R in terms of M and v in terms of t, and recalling the relation between the vectorization of an operator and its Bloch vector in Equation (2.11), we have that (denoting $s = \mathrm{bvec}(\Lambda(X))$ and $r = \mathrm{bvec}(X)$)

$$\mathrm{vec}(\Lambda(X)) = S_\Lambda \mathrm{vec}(X) \tag{2.16}$$

$$\begin{pmatrix} \sqrt{1/m} \\ \sqrt{\frac{(m-1)}{m}} s \end{pmatrix} = \left(\begin{array}{c|c} \sqrt{n/m} & 0^T \\ \hline \sqrt{n}\sqrt{\frac{m-1}{m}} t & (\sqrt{\frac{m-1}{m}}/\sqrt{\frac{n-1}{n}}) M \end{array}\right) \begin{pmatrix} \sqrt{1/n} \\ \sqrt{\frac{(n-1)}{n}} r \end{pmatrix} \tag{2.17}$$

$$= \begin{pmatrix} \sqrt{1/m} \\ \sqrt{\frac{(m-1)}{m}} (Mr + t) \end{pmatrix}, \tag{2.18}$$

verifying our expressions for M and t.

Note that the Bloch representation of a linear combination of channels is equal to the linear combination of the Bloch representations of the originals. That is, for $\Lambda_1, \Lambda_2 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, with Bloch representations (M_1, t_1) and (M_2, t_2), the Block representation of $\alpha\Lambda_1 + \beta\Lambda_2$ is $(\alpha M_1 + \beta M_2, \alpha t_1 + \beta t_2)$. This follows from the linearity of the bvec operation.

2.3 The Process χ-Matrix, Average Gate Fidelity, and Pauli twirling

Any arbitrary map $\Lambda \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ can be written as $\Lambda(X) = \sum_{j=1}^{k} A_j X A_j^\dagger$, for some operators $\{A_j : j = 1, ..., n\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$. Given a basis $\{P_j\}_{j=1}^{\dim(\mathcal{X})\dim(\mathcal{Y})}$ for $\mathrm{L}(\mathcal{X}, \mathcal{Y})$, one can expand $A_i = \sum_j \alpha_{ij} P_j$ for some complex numbers α_{ij}. In this way one can write $\Lambda(X) = \sum_{ijk} \alpha_{ij}\alpha_{ik}^* P_j X P_k^\dagger = \sum_{jk} \chi_{jk} P_j X P_k^\dagger$, where $\chi_{jk} = \sum_i \alpha_{ij}\alpha_{ik}^*$ is called the process matrix [16].

The *average gate fidelity* of a quantum process $\Lambda \in \mathrm{C}(\mathcal{X})$, which is the imperfect implementation of some unitary U, is defined as

$$F(\Lambda, U) = \int d\psi \, \langle\psi| U^\dagger \Lambda(|\psi\rangle\langle\psi|) U |\psi\rangle .$$

In [17], it was shown that

$$F(\Lambda, U) = \frac{\sum_k |\mathrm{Tr}(A_k)|^2 + d}{d^2 + d},$$

where $d = \dim(\mathcal{X})$, and $\{A_k\}_k \subset \mathrm{L}(\mathcal{X})$ is a Kraus representation for $U^\dagger \circ \Lambda$ or $\Lambda \circ U^\dagger$. Now, as Λ is an erroneous implementation of U, $\mathcal{E} = \Lambda \circ U^\dagger$ is the error on the gate U for the implementation Λ. If one uses a Bloch basis to expand the error $\mathcal{E}$, then $|\mathrm{Tr}(A_k)|^2 = |\sum_j \alpha_{kj}\mathrm{Tr}(P_j)|^2 = |\alpha_{k0}\sqrt{d}|^2$, and thus $\sum_k |\mathrm{Tr}(A_k)|^2 = \sum_k d|\alpha_{k0}|^2 = \chi_{00}$. Thus, it follows that

$$F(\Lambda, U) = \frac{\chi_{00} + 1}{d + 1},$$

where χ is the the process matrix for $\Lambda \circ U^\dagger$ in any Bloch basis.

For a Pauli channel $\mathcal{E} \in \mathrm{C}(\mathcal{X} = \mathbb{C}^{2^n})$, we can write its χ-matrix in terms of the normalized basis of tensor product Pauli operators for $\mathrm{L}(\mathcal{X})$. In this case, χ is diagonal, with $\chi_{jj} = dp_j$, where $d = 2^n$ and p_j is the probability that the j^{th} Pauli operator (according to whatever ordering is chosen) is applied. For an arbitrary map Λ, the effect of Pauli twirling can be computed easily by simply removing the off-diagonal elements of the χ-matrix for Λ in the Pauli basis [18].

Chapter 3

Quantum Information Background

3.1 Error and Distinguishability

When one wishes to quantify how bad an error messes up a quantum state, it is necessary to quantify how different (or similar) two quantum states are. Of course, quantum states are represented as density operators, and there is no shortage of norms on the space of operators from which distance measures between states can be induced. In particular, we choose to use the Schatten p-norms (see Section 2.1.5) to quantify the difference between states. For any $p \geq 1$ and $\rho \in \mathrm{D}(\mathcal{X})$, the amount that an "error" $\Lambda \in \mathrm{C}(\mathcal{X})$ "messes up" ρ can be defined as $\|\rho - \Lambda(\rho)\|_p$; that is, how far does Λ move the state according to $\|\cdot\|_p$?

While any choice of $p \geq 1$ is fine in principle, it is highly desirable to have an *operationally motivated* way of quantifying how different two quantum states are. A strong way of defining how operationally different two states are, is by trying to quantify how hard it would be tell them apart. That is, if you are given one of two states $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ (but you don't know which one), and you can do whatever you want with it, how likely are you to be able to determine which state you were given?

The above scenario can be formalized in the following game. Let $p : \{0,1\} \to [0,1]$ be a probability distribution (that is, $p(0) = \lambda$ and $p(1) = 1 - \lambda$ for some $\lambda \in [0,1]$), and let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be two states. The parameter λ and states ρ_0 and ρ_1 are known to the player. A bit $\alpha \in \{0,1\}$ is sampled according to p, and a single copy of ρ_α is given to the player. The player is allowed to do anything with ρ_α, and they win if they can successfully guess α.

In this scenario, effectively all the player can do is measure the state, as preceding a measurement by some CPTP map is equivalent to making a different measurement. Also,

the player might as well make only a two outcome measurement, where each outcome is labelled by whether or not the player guesses 0 or 1. Thus, given a particular measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$ ($\mu(0) + \mu(1) = \mathbb{1}_{\mathcal{X}}$), the probability that the player correctly guess the sampled bit is given by $\lambda\langle\mu(0), \rho_0\rangle + (1-\lambda)\langle\mu(1), \rho_1\rangle$. We're of course interested in the optimal success probability of this game, which is the maximization of this quantity over all possible measurements. This brings us to what is referred to as the Holevo-Helstrom Theorem [19].

Theorem 3.1.1. *For $\rho_0, \rho_1 \in D(\mathcal{X})$, and $\lambda \in [0,1]$, and $\mu : \{0,1\} \to Pos(\mathcal{X})$ a measurement,*

$$\lambda\langle\mu(0), \rho_0\rangle + (1-\lambda)\langle\mu(1), \rho_1\rangle \leq \frac{1}{2} + \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1. \tag{3.1}$$

There exists a projective measurement that achieves equality.[1]

This theorem tells us that the optimal success probability of the above game is given by $\frac{1}{2} + \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1$. Using this, the *distinguishability* of two quantum states ρ_0 and ρ_1 is defined as $\frac{1}{2} + \frac{1}{4}\|\rho_0 - \rho_1\|_1$, or equivalently, as the optimal success probability of winning the above game when $\lambda = \frac{1}{2}$. Thus, discarding the pre-factors, $\|\rho_0 - \rho_1\|_1$ is an excellent distance measure for quantifying how different two quantum states are.

One can further extend the above game to ask about the distinguishability between two channels $\Lambda_0, \Lambda_1 \in \mathrm{C}(\mathcal{X})$. Again, let $p : \{0,1\} \to [0,1]$ be a probability distribution (where $p(0) = \lambda$, and $p(1) = 1 - \lambda$ for some $\lambda \in [0,1]$). We can ask what is the optimal success probability of winning the following game. Let $\alpha \in \{0,1\}$ be sampled according to p. Given a single use of Λ_α, the player must guess what α is. This game can be essentially reduced to the same game as before, by noting that once one uses Λ_α once (by passing a state through it), the problem reduces to trying to distinguish the outputs of Λ_0 and Λ_1. One subtlety however arises in that, for some complex Euclidean space $\mathcal{Z}$, one could choose to pass a state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ through $\Lambda_\alpha \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}$, possibly utilizing entanglement to get an advantage. It turns out that, in general, there is an optimal $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ when $\dim(\mathcal{Z}) = \dim(\mathcal{X})$ (for some cases, it isn't necessary for $\mathcal{Z}$ to be as large as $\mathcal{X}$). For a given choice of $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ and measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$ (where α is guessed if $\mu(\alpha)$ is observed), the probability that the player wins the game is given as $\lambda\langle\mu(0), (\Lambda_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\rangle + (1-\lambda)\langle\mu(1), (\Lambda_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)\rangle$. The optimal success probability of this game is provided by the following theorem, in analogy to the state distinguishability game [13].

[1] An example of an optimal projective measurement is given by $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$, where $\mu(0)$ is the orthogonal projector onto the span of eigenvectors with positive eigenvalues of the operator $\lambda\rho_0 - (1-\lambda)\rho_1$, and $\mu(1) = \mathbb{1}_{\mathcal{X}} - \mu(0)$ [13].

Theorem 3.1.2. *Let $\mathcal{X}$ and $\mathcal{Z}$ be complex Euclidean spaces, $\Lambda_0, \Lambda_1 \in C(\mathcal{X})$, $\rho \in D(\mathcal{X} \otimes \mathcal{Z})$, $\mu : \{0,1\} \to Pos(\mathcal{X} \otimes \mathcal{Z})$ a measurement, and let $\lambda \in [0,1]$. It follows that*

$$\lambda \langle \mu(0), (\Lambda_0 \otimes \mathbb{1}_{L(\mathcal{Z})})(\rho) \rangle + (1-\lambda) \langle \mu(1), (\Lambda_1 \otimes \mathbb{1}_{L(\mathcal{Z})})(\rho) \rangle \leq \frac{1}{2} + \frac{1}{2} \| \lambda \Lambda_0 - (1-\lambda)\Lambda_1 \|_\diamond, \tag{3.2}$$

with equality achieved for some choice of ρ and μ when $dim(\mathcal{Z}) = dim(\mathcal{X})$.

In the above theorem, the function $\| \cdot \|_\diamond$ is used. This function, commonly referred to as the "diamond norm" (it is in fact a norm) is defined in the following way [20] (note that this is just one of several equivalent definitions). For $\Lambda \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$,

$$\|\Lambda\|_\diamond = \sup\{ \|(\Lambda \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(X)\|_1 : X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y}), \|X\|_1 \leq 1 \}. \tag{3.3}$$

When $\lambda = 1/2$, the optimal success probability of distinguishing $\Lambda_0, \Lambda_1 \in \mathrm{C}(\mathcal{X})$ is given as $\frac{1}{2} + \frac{1}{4}\|\Lambda_0 - \Lambda_1\|_\diamond$. Thus, discarding the pre-factors, $\|\Lambda_0 - \Lambda_1\|_\diamond$ is a well motived quantity for measuring how different the two channels are. This will be useful when we are interested in finding approximations to channels.

3.2 Efficiently Simulable Sub-theories

Quantum mechanics is not efficiently simulable on a classical computer in general, as the number of parameters required to describe a system grows exponentially as the number of parts of the system grows. Nevertheless, if one restricts the initial states and operations allowable by quantum theory, in some cases, this can lead to an efficiently simulable *sub-theory* of quantum theory. This can be understood intuitively by simply noting that if one limits the number of degrees of freedom accessible to a system (by wisely choosing the types of states and operations you wish to investigate), then the number of parameters required to describe the system can be reduced from exponential to polynomial in the number of subsystems, thereby admitting a compact description of the state of the system.

The most well known and commonly used efficiently simulable sub-theory is provided by the Gottesman-Knill (GK) theorem [21]. The theorem, by means of an explicit simulation algorithm, gives that perfect "Clifford" operations and Pauli-basis measurements ("stabilizer circuit elements") are efficiently simulable on "stabilizer states". We omit a detailed description of this sub-theory, as the details of it are not important for this thesis, but the basic idea is as follows. A "stabilizer" state on n-qubits is defined as the unique

+1 eigenvector (of unit length, ignoring global phase) of n, mutually commuting Pauli operators (called the *stabilizer*). Within this context, the Clifford operators are defined as unitaries which, under the action of conjugation, permute the Pauli group. Thus, the action of a Clifford on a stabilizer state can be tracked by acting the Clifford on the stabilizer to get a new one, which stabilizes the output of the Clifford on the original state (and therefore uniquely specifies it). The action of Pauli measurements is similar, though indeterminacy must be handled.

The GK algorithm can be extended to simulate the action of certain types of noisy circuits on stabilizer states. Specifically, if errors are modelled as probabilistic application of stabilizer circuit elements (Clifford gates and Pauli measurements), then one can simulate the action of a noisy stabilizer circuit on a stabilizer state via Monte Carlo (MC). The combination of these techniques (GK-MC) has been used to study performance of error correcting circuits subject to limited error models. Examples include the numerical estimation of threshold error rates in topological codes [1, 7, 8], and, in conjunction with Sequential Monte-Carlo techniques, has enabled reasoning about overhead required even when well below threshold [10].

While GK-MC is currently the most prominent method for efficiently simulating noisy circuits (not least of all because of its applicability to one of the most important classes of QEC protocols), and is therefore the example of choice within this thesis, we note that the work in this thesis is not tied to GK-MC, and is applicable within the context of other efficiently simulable sub-theories. These include Wigner function simulation [22], matchcircuit simulation [23], quantum normalizer circuit simulation [24, 25], or the non-adaptive strong-simulation algorithm for Clifford circuits [26].

Chapter 4

Partial Order on Quantum Channels and Honest Approximations

We now formally define the notion of one map being an honest representation of another. We start by defining a family of partial orders on $\mathrm{C}(\mathcal{X})$ which encapsulate possible definitions for a channel $\Lambda \in \mathrm{C}(\mathcal{X} = \mathbb{C}^n)$ to be somehow *globally* worse than another channel $\mathcal{E} \in \mathrm{C}(\mathcal{X})$, and describe results that give sufficient conditions for when some of the partial ordering relations hold for two maps. We leave the proofs of these results to Appendix B, as they are technical and straightforward.

Using these partial order results, we define an optimization problem for finding honest approximations to arbitrary error maps, and apply them to some simple examples for intuition.

4.1 Partial Ordering Results

Definition 4.1.1. For channels $\Lambda, \Gamma \in \mathrm{C}(\mathcal{X})$, and $1 \leq p \leq \infty$, we write $\Lambda \geq_p \Gamma$ if

$$\|\rho - \Lambda(\rho)\|_p \geq \|\rho - \Gamma(\rho)\|_p \tag{4.1}$$

for all $\rho \in \mathrm{D}(\mathcal{X})$. We write $\Lambda \geq_p^* \Gamma$ to denote the restriction of this statement to pure states. If $\Lambda \geq_p \Gamma$ ($\Lambda \geq_p^* \Gamma$) and $\Gamma \geq_p \Lambda$ ($\Gamma \geq_p^* \Lambda$), then we write $\Lambda \sim_p \Gamma$ ($\Lambda \sim_p^* \Gamma$).

The relation $\geq_p$ ($\geq_p^*$) is a *partial order* on $\mathrm{C}(\mathcal{X})$.

The case we are most interested in is when $p = 1$, due to the relationship between $\|\cdot\|_1$ and distinguishability. Nevertheless, we will say that Λ is an *honest approximation* of $\mathcal{E}$ if $\Lambda \geq_p \mathcal{E}$ (or $\Lambda \geq_p^* \mathcal{E}$) for some $1 \leq p \leq \infty$. Note that this definition differs from [11, 12], where we explicitly reserve the designation "honest" for the case when $p = 1$. We redefine it here to avoid having to constantly qualify statements about honest approximations, as what p is meant is always clear within context. Below we add one more partial order to the mix, involving the Fidelity function (see Section 2.1.6).

Definition 4.1.2. For channels $\Lambda, \Gamma \in \mathrm{C}(\mathcal{X})$, we write $\Lambda \geq_F \Gamma$ if

$$F(\Lambda(\rho), \rho) \leq F(\Gamma(\rho), \rho)$$

for all $\rho \in \mathrm{D}(\mathcal{X})$. We write $\Lambda \geq_F^* \Gamma$ to denote the restriction of the above statement to pure states.

The following results provide sufficient conditions for a channel to be an honest representation of another. They are corollaries to statements proven in Appendix B.

Corollary 4.1.1. *Let $\Lambda, \Gamma \in C(\mathcal{X} = \mathbb{C}^n)$ and let their Bloch representations be denoted (M_Λ, t_Λ) and (M_Γ, t_Γ) respectively. Define*

$$A = (\mathbb{1}_{n^2-1} - M_\Lambda)^T(\mathbb{1}_{n^2-1} - M_\Lambda) - (\mathbb{1}_{n^2-1} - M_\Gamma)^T(\mathbb{1}_{n^2-1} - M_\Gamma) \tag{4.2}$$

$$+ (\|t_\Lambda\|_2^2 - \|t_\Gamma\|_2^2 - 2\|(\mathbb{1}_{n^2-1} - M_\Lambda)^T t_\Lambda - (\mathbb{1}_{n^2-1} - M_\Gamma)^T t_\Gamma\|_2)\mathbb{1}_{n^2-1}, \tag{4.3}$$

Where $\mathbb{1}_{n^2-1} \in \mathbb{R}^{(n^2-1)\times(n^2-1)}$ is the identity matrix. It holds that if $A \geq 0$, then $\Lambda \geq_2^* \Gamma$.

If, in addition,

$$\|t_\Lambda\|_2^2 - \|t_\Gamma\|_2^2 = 0 \tag{4.4}$$

$$\|(\mathbb{1}_{n^2-1} - M_\Lambda)^T t_\Lambda - (\mathbb{1}_{n^2-1} - M_\Gamma)^T t_\Gamma\|_2 = 0, \tag{4.5}$$

then $\Lambda \geq_2 \Gamma$.

Corollary 4.1.2. *For $\Lambda, \Gamma \in C(\mathcal{X} = \mathbb{C}^n)$ and $p \geq 1$, if $\mathrm{rank}(\rho - \Gamma(\rho)) \leq 2$ for all pure states $\rho \in D(\mathcal{X})$ and $\Lambda \geq_p^* \Gamma$, then $\Lambda \geq_1^* \Gamma$.*

Corollary 4.1.3. *Let $\Lambda, \Gamma \in C(\mathcal{X} = \mathbb{C}^2)$. For any $q, p \geq 1$, $\Lambda \geq_q \Gamma \Leftrightarrow \Lambda \geq_p \Gamma$ ($\Lambda \geq_q^* \Gamma \Leftrightarrow \Lambda \geq_p^* \Gamma$).*

Corollary 4.1.4. *Let $\Lambda, \Gamma \in C(\mathcal{X} = \mathbb{C}^n)$, $A \in L(\mathcal{X}, \mathcal{X} \otimes \mathcal{Z})$ be a Stinespring representation of Γ for some space $\mathcal{Z}$, $u \in \mathcal{Z}$ an arbitrary unit vector, and (M_Λ, t_Λ) the Bloch representation of Λ. Define the maps $\Gamma_1, \Gamma_2 \in C(\mathcal{X}, \mathcal{X} \otimes \mathcal{Z})$ as $\Gamma_1(X) = X \otimes uu^\dagger$ and $\Gamma_2(X) = AXA^\dagger$ for $X \in L(\mathcal{X})$, and let $(M_{\Gamma_1}, t_{\Gamma_1})$ and $(M_{\Gamma_2}, t_{\Gamma_2})$ be their respective Bloch representations. Define*

$$\begin{aligned} B = &(\mathbb{1}_{n^2-1} - M_\Lambda)^T(\mathbb{1}_{n^2-1} - M_\Lambda) - (M_{\Gamma_1} - M_{\Gamma_2})^T(M_{\Gamma_1} - M_{\Gamma_2}) \\ &+ (\|t_\Lambda\|_2^2 - \|t_{\Gamma_1} - t_{\Gamma_2}\|_2^2 \\ &- 2\|(\mathbb{1}_{n^2-1} - M_\Lambda)^T(-t_\Lambda) - (M_{\Gamma_1} - M_{\Gamma_2})^T(t_{\Gamma_1} - t_{\Gamma_2})\|_2)\mathbb{1}_{n^2-1}. \end{aligned}$$

It holds that if $B \geq 0$, then $\Lambda \geq_1^* \Gamma$ and $\Gamma \geq_F^* \Lambda$.

Corollary 4.1.1 is a slightly generalized version of the result in [11], which takes into account the possible non-unitality of the map Λ. This condition is used throughout the rest of this thesis as the condition in an optimization problem for finding honest approximations in the $p = 2$ sense. As already noted, the case we're most interested in is honesty in the $p = 1$ sense. Corollary 4.1.2, gives that honesty in the $p = 2$ sense is sufficient for the $p = 1$ sense if $\text{rank}(\rho - \mathcal{E}(\rho)) \leq 2$ for all pure states ρ (and, in particular, for all qubit maps). As an example, if $\mathcal{E}$ is a unitary map, this condition is satisfied. Finally, Corollary 4.1.4 provides a condition that is sufficient to ensure honesty in the $p = 1$ sense for arbitrary maps in arbitrary dimensions, and with an added bonus is also sufficient for the Fidelity partial order to hold.

4.2 Approximation Algorithm

We now use Corollary 4.1.1 to generate honest approximations to a given channel $\mathcal{E} \in \text{C}(\mathcal{X})$. As we're interested in classically simulable approximations, we consider optimization problems over convex combinations of some specified set of channels. Eg. for GK-MC simulable channels, this would be some subset of channels achievable using only Clifford operations and Pauli measurements. Thus, we have the following optimization problem (as given in [12]):

$$
\begin{aligned}
&\text{input: Finite set of channels } \{\Lambda_i\}_{i=1}^n \text{and channel } \mathcal{E} \\
&\text{minimize: } f(p_1, ..., p_n) = \left\| \sum_{i=1}^{n} p_i \Lambda_i - \mathcal{E} \right\|_\diamond \\
&\text{s.t. } \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \text{ and } A \geq 0, \text{ where} \\
&\quad A = (\mathbb{1} - M_\Lambda)^T (\mathbb{1} - M_\Lambda) - (\mathbb{1} - M_\mathcal{E})^T (\mathbb{1} - M_\mathcal{E}) + \\
&\quad (\|\vec{t}_\Lambda\|_2^2 - \|\vec{t}_\mathcal{E}\|_2^2 - 2\|(\mathbb{1} - M_\Lambda)^T \vec{t}_\Lambda - (\mathbb{1} - M_\mathcal{E})^T \vec{t}_\mathcal{E}\|_2) \mathbb{1},
\end{aligned}
\tag{4.6}
$$

where $M_\Lambda = \sum_{i=1}^n p_i M_i$, $\vec{t}_\Lambda = \sum_{i=1}^n p_i \vec{t}_i$, $(M_i, \vec{t}_i)$ is the Bloch representation of channel Λ_i, $(M_\mathcal{E}, \vec{t}_\mathcal{E})$ is the Bloch representation of $\mathcal{E}$, and $\mathbb{1}$ is the identity matrix of appropriate size.[1]

We implement the approximation optimization problem in MATLAB using the built in fmincon function. The diamond norm is computed using a semi-definite program given by Watrous [27] and implemented using the CVX package [28]. Linear constraints ensure that the vector $(p_1, ..., p_n)$ is a probability vector and non-linear constraints check that the eigenvalues of the matrix A are non-negative with a tolerance of 10^{-15}. The SQP algorithm is used for the optimization. Due to the non-convexity of this problem it is necessary to run many local solvers and then choose the best result. This is done using the MultiStart function which instantiates the local solver many times over randomly chosen starting points that satisfy the constraints. For every approximation in this thesis we used 72 starting points (a common, high-ish multiple of the number of cores available on the computers used for this work).

4.3 Metrics for Judging Approximations

Given a set of approximations for some error $\mathcal{E}$, we need some metrics for checking how honest they are, and comparing how they perform next to each other. The metrics given here are taken from [12] with minimal modification.

[1] The complexity of this optimization problem has not been analyzed, but it certainly doesn't scale well with the dimension or with the number of input channels used to find an approximation. For example, when the system is a qubit and the input set is the set of Pauli operators, one run of our implementation takes less than a minute. When the system is two qubits (with the input set being two qubit Pauli operators), it takes on the order of an hour. Finding a better algorithm has, however, not been a priority.

As, in the end, we're interested in talking about errors on *gates*, we set some notation consistent with that. Suppose $\Lambda_{\text{Original}}$ is a map representing an imperfect implementation of some ideal operation U_{Ideal}. Then, we can write $\Lambda_{\text{Original}} = \mathcal{E} \circ U_{\text{Ideal}}$, where $\mathcal{E}$ represents the error. Given an approximation $\Lambda_{\mathcal{E}}$ to the error $\mathcal{E}$, we let the operation $\Lambda = \Lambda_{\mathcal{E}} \circ U_{\text{Ideal}}$ denote the approximation to the gate $\Lambda_{\text{Original}}$.

Several metrics are used to compare how each approximation performs on individual gates. In what follows we denote the noisy implementation of some ideal operation U_{ideal} by $\Lambda_{\text{Original}}$, and use Λ as a place holder for the various approximations. The first few metrics are well known quantities.

- χ_{00} — The first entry of the χ-process matrix of the error in the Pauli basis [16]. This quantity is reported due to its relation to the average gate fidelity (see [17] and Section 2.3), and, for Pauli channels, is the probability that the identity operation occurs.
- $\|\Lambda - U_{\text{Ideal}}\|_\diamond$ and $\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$ — The distance of the approximation Λ to the ideal gate and original error, respectively.

The rest stem from our definition of honesty. Using the function

$$h(\Gamma, \mathcal{E}, \rho) \equiv \|\rho - \Gamma(\rho)\|_1 - \|\rho - \mathcal{E}(\rho)\|_1, \tag{4.7}$$

which we call the *hedging* of the channel Γ relative to $\mathcal{E}$ for the state ρ, the statement that Λ honestly represents the error of $\Lambda_{\text{Original}}$ can be restated as $h(\Lambda, \Lambda_{\text{Original}}, \rho) \geq 0$ for all pure states ρ. We calculate three quantities related to the hedging, which we approximate by randomly sampling N pure states $\{|\psi_i\rangle\}_{i=1}^N$ according to the Haar measure.

- $\bar{h}(\Lambda, \Lambda_{\text{Original}}) \equiv \int d\psi \, h(\Lambda, \Lambda_{\text{Original}}, |\psi\rangle\langle\psi|)$
  $\approx \frac{1}{N} \sum_{i=1}^N h(\Lambda, \Lambda_{\text{Original}}, |\psi_i\rangle\langle\psi_i|)$ — The Haar average of the hedging function over pure states.
- $p_{\text{viol}} \approx \frac{N_{\text{viol}}}{N}$ — The number of sampled pure states $|\psi\rangle$ for which honesty is violated (that is, $h(\Lambda, \Lambda_{\text{Original}}, |\psi\rangle\langle\psi|) < 0$) divided by the total number, where $N_{\text{viol}} = |\{|\psi\rangle \in \{|\psi_i\rangle\}_{i=1}^N : h(\Lambda, \Lambda_{\text{Original}}, |\psi\rangle\langle\psi|) < 0\}|$.

Throughout this thesis we compute the latter two metrics by sampling $N = 10^6$. In Appendix C we give a quick calculation and argument to show that $N = 10^6$ provides a very good estimate of $\bar{h}$ in the vast majority of cases in this thesis, and most importantly, in Table 5.2 in Section 5.5, where the usefulness of our method for bounding circuit performance is tested. We also show that the estimate of p_{viol} (the probability that a randomly chosen pure state has negative hedging) is very good in all cases.

4.4 Instructive Qubit Examples

We now apply this approximation algorithm to some instructive qubit examples, where we can use the Bloch sphere to visualize how the channels warp state space. We start with approximations to small unitary rotations, following examples given in [11], and end with approximations to a (non-unital) depolarizing channel. In both cases, we approximate the errors using Pauli channels, and add some additional, efficiently-simulable-by-GK-MC (see Section 3.2), operations to attain better representation of the errors. Note that the non-unital examples are similar to those in [29].

4.4.1 Unitary Rotations

Consider the three unitary rotations, $\Lambda_k(\rho) = U_k \rho U_k^\dagger$, where $U_k = \exp\left(-i\frac{\theta}{2}\vec{n}_k \cdot \vec{\sigma}\right)$, $\vec{n}_k = (\sin(k\pi/8), 0, \cos(k\pi/8))$, and $\vec{\sigma} = (X, Y, Z)$, for $k \in \{0, 1, 2\}$, and $\theta = 0.02$. We approximate these channels as a Pauli channel, a mixed-Clifford channel (where we include all single qubit Cliffords), and the Pauli twirled channel. Statistics for each approximation are given in Table 4.1, where $N = 10^6$ pure states were used to generate the hedging statistics. Note that, in principle, the statistics for the Clifford approximations of Λ_0 and Λ_2 should be the same. The Clifford approximation for Λ_0 is a probabilistic application a $\pi/2$ rotation about the z-axis. The analogue for this in the Λ_2 case would be a probabilistic application of the Hadamard gate, which, by symmetry, would have the same hedging statistics and distance to the Original as the Clifford approximation has for Λ_0 (which would be closer, and therefore a better approximation). This is an example of how the approximation algorithm is not particularly well behaved.

The effect of the unitary rotations, the Pauli approximations, and Pauli twirled approximations, on the x-z axis of the Bloch sphere are plotted in Figure 4.1 (where we instead use the rotation angle $\theta = 2\sin^{-1}(\sqrt{0.1}) \approx 0.64$ to exaggerate the visual features of the approximations). The action of the unitary rotation, which rotates states out of the x-z plane, is represented as a dephasing channel about the rotation axis. Using Λ_0 as an example, this is done by noting that, for any qubit state ρ, $\|\Lambda_0(\rho) - \rho\|_1 = 2\left|\sin\left(\theta/2\right)\right| \left|\sin(\alpha)\right| \|r\|_2$, where r is the Bloch vector of ρ, and α is the angle of r away from the z axis. If one considers a dephasing channel about z, given by $\Lambda(\rho) = (1-p)\rho + pZ\rho Z$, one sees that for any qubit state ρ, $\|\Lambda(\rho) - \rho\|_1 = 2p\left|\sin(\alpha)\right| \|r\|_2$, where α and r are defined in the same way as previously. Thus, by setting $p = \left|\sin(\theta/2)\right|$ for the dephasing channel, it can *exactly* reproduce the input-output distinguishability profile of the rotation. It is this dephasing channel that we plot in Figure 4.1 in place of each unitary, for easy visual comparison to

the approximations. Notice that for Λ_0, the average hedging of both the Pauli and Clifford approximations is 0, and visually, in Figure 4.1, one can confirm that the honest Pauli approximation is exactly the dephasing channel that perfectly reproduces the input-output distinguishability of Λ_0.

Table 4.1: Statistics for approximations to the unitary rotations for $\theta = 0.02$ (which yields $\chi_{00} \approx 0.9999$). $N = 10^6$ pure states were sampled for the hedging statistics.

Unitary Example					
Gate	Statistics	Original	Pauli	Clifford	Pauli twirled
Λ_0	χ_{00}	0.999900	0.990000	0.992929	0.999900
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	2.00×10^{-2}	2.00×10^{-2}	2.00×10^{-2}	2.00×10^{-4}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.81×10^{-2}	1.51×10^{-2}	2.00×10^{-2}
	$\overline{h}$		0.	0.	-1.56×10^{-2}
	p_{viol}		0.	0.	1.
Λ_1	χ_{00}	0.999900	0.986025	0.990489	0.999900
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	2.00×10^{-2}	2.80×10^{-2}	2.34×10^{-2}	2.00×10^{-4}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		3.59×10^{-2}	2.05×10^{-2}	2.00×10^{-2}
	$\overline{h}$		3.29×10^{-3}	8.15×10^{-4}	-1.56×10^{-2}
	p_{viol}		0.	0.	1.
Λ_2	χ_{00}	0.999900	0.985000	0.989794	0.999900
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	2.00×10^{-2}	3.00×10^{-2}	2.41×10^{-2}	2.00×10^{-4}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		3.81×10^{-2}	2.23×10^{-2}	2.00×10^{-2}
	$\overline{h}$		4.29×10^{-3}	9.76×10^{-4}	-1.56×10^{-2}
	p_{viol}		0.	0.	0.99999

4.4.2 Non-unital Channels

Let $\mathcal{E}(X) = K_1 X K_1^\dagger + K_2 X K_2^\dagger$ for

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \; K_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \tag{4.8}$$

Define $\mathcal{E}_\alpha = U_\alpha \circ \mathcal{E} \circ U_\alpha^\dagger$, for $\alpha \in \{0, 1, 2\}$ and $U_\alpha = \exp\left(-i\frac{\alpha\pi}{8}Y\right)$. These channels are examples of amplitude damping channels which drive the system towards a particular state.

We approximate these channels as a Pauli channel, something we label as a Pauli & Measurement channel, and a Pauli twirled channel. A Pauli & Measurement approximation here denotes an approximation which uses the following channels:

$$\{\mathbb{1}, X, Y, Z, \Phi_{x+}, \Phi_{x-}, \Phi_{y+}, \Phi_{y-}, \Phi_{z+}, \Phi_{z-}\},$$

where $\Phi_{\alpha\pm}(X) = |\alpha\pm\rangle\langle\alpha\pm|$ for $\alpha \in \{X, Y, Z\}$, and $|\alpha\pm\rangle$ satisfies $\alpha\,|\alpha\pm\rangle = \pm\,|\alpha\pm\rangle$. (In the above list, $\{\mathbb{1}, X, Y, Z\}$ denote both the Pauli operators themselves, as well as the channels defined by conjugation of these unitaries.) For example Φ_{x+} maps every state to the $+1$ eigenstate of the X operator. This channel can be viewed as measuring the state in the X basis, then conditionally performing a Y or Z gate on the state if -1 is measured. As these channels consist of Pauli measurements and conditional Pauli gates, they are efficiently simulable. Channels like this enable the generation of efficiently simulable non-unital gates [29].

Statistics for these approximations are given in Table 4.2, and the visualizations for the Pauli approximations are given in Figure 4.2, and for the Pauli & Measurement in Figure 4.3 (for the visualizations, we set $\gamma = 0.25$ to exaggerate the effects on the Bloch sphere). One sees that, as expected, as Pauli & Measurement channels can generate non-unital effects, they do a much better job of approximating the various amplitude damping channels. Notice that for $\mathcal{E}_0$, the Pauli & Measurement approximation is able to "touch" the top of the deformed Bloch sphere of $\mathcal{E}_0$, whereas this is not the case for $\mathcal{E}_1$ or $\mathcal{E}_2$. The reason for this is that the "Measurement" part of Pauli & Measurement channels can only prepare states that are convex combinations of Pauli eigenstates, and as such, when the amplitude damping channel is driving the system to a Pauli eigenstate, the Pauli & Measurement channel can approximate it quite closely. When the amplitude damping channel isn't driving to a convex combination of Pauli eigenstates, as for $\mathcal{E}_1$ and $\mathcal{E}_2$, the Pauli & Measurement channels can't approximate it quite as well, as they can't push the centre of the Bloch sphere outside of the convex set. It is also clear that the Pauli & Measurement channels provide a much better fit to the I/O Distinguishability profiles of the Original.

4.4.3 Foreshadowing Observations

Before moving on, it is worth noting a couple of observations in anticipation of what will follow. The parameters in the unitary and non-unital examples were chosen so that the Original channels would all have the same χ_{00} element (equivalently, the same average gate fidelity). In the case of the unitary examples, we see that the χ_{00} element of the honest

Table 4.2: Statistics for approximations to the amplitude damping channels for $\gamma = 2\times10^{-4}$ (which yields $\chi_{00} \approx 0.9999$). $N = 10^6$ pure states were sampled for the hedging statistics.

Amplitude Damping					
Gate	Statistics	Original	Pauli	Pauli & Measurement	Pauli Twirled
$\mathcal{E}_0$	χ_{00}	0.999900	0.999720	0.999850	0.999900
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.00×10^{-4}	5.61×10^{-4}	4.00×10^{-4}	2.00×10^{-4}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.23×10^{-4}	1.00×10^{-4}	2.00×10^{-4}
	$\overline{h}$		1.53×10^{-4}	4.60×10^{-5}	-8.26×10^{-5}
	p_{viol}		0.	0.	0.74941
$\mathcal{E}_1$	χ_{00}	0.999900	0.999713	0.999830	0.999900
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.00×10^{-4}	5.75×10^{-4}	4.06×10^{-4}	2.00×10^{-4}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.33×10^{-4}	1.41×10^{-4}	2.35×10^{-4}
	$\overline{h}$		1.63×10^{-4}	5.03×10^{-5}	-8.43×10^{-5}
	p_{viol}		0.	0.	0.75218
$\mathcal{E}_2$	χ_{00}	0.999900	0.999710	0.999824	0.999900
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.00×10^{-4}	5.80×10^{-4}	4.10×10^{-4}	2.00×10^{-4}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.37×10^{-4}	1.54×10^{-4}	2.50×10^{-4}
	$\overline{h}$		1.66×10^{-4}	5.27×10^{-5}	-8.59×10^{-5}
	p_{viol}		0.	0.	0.74498

approximations is much lower than that of the Original as compared to the non-unital example. Also, in the unitary example, the hedging of the Pauli Twirled approximation is much worse than in the non-unital example. This observation, and the contrast between the performance of the honest approximations and Pauli Twirled channels is explained in detail in Chapter 6.

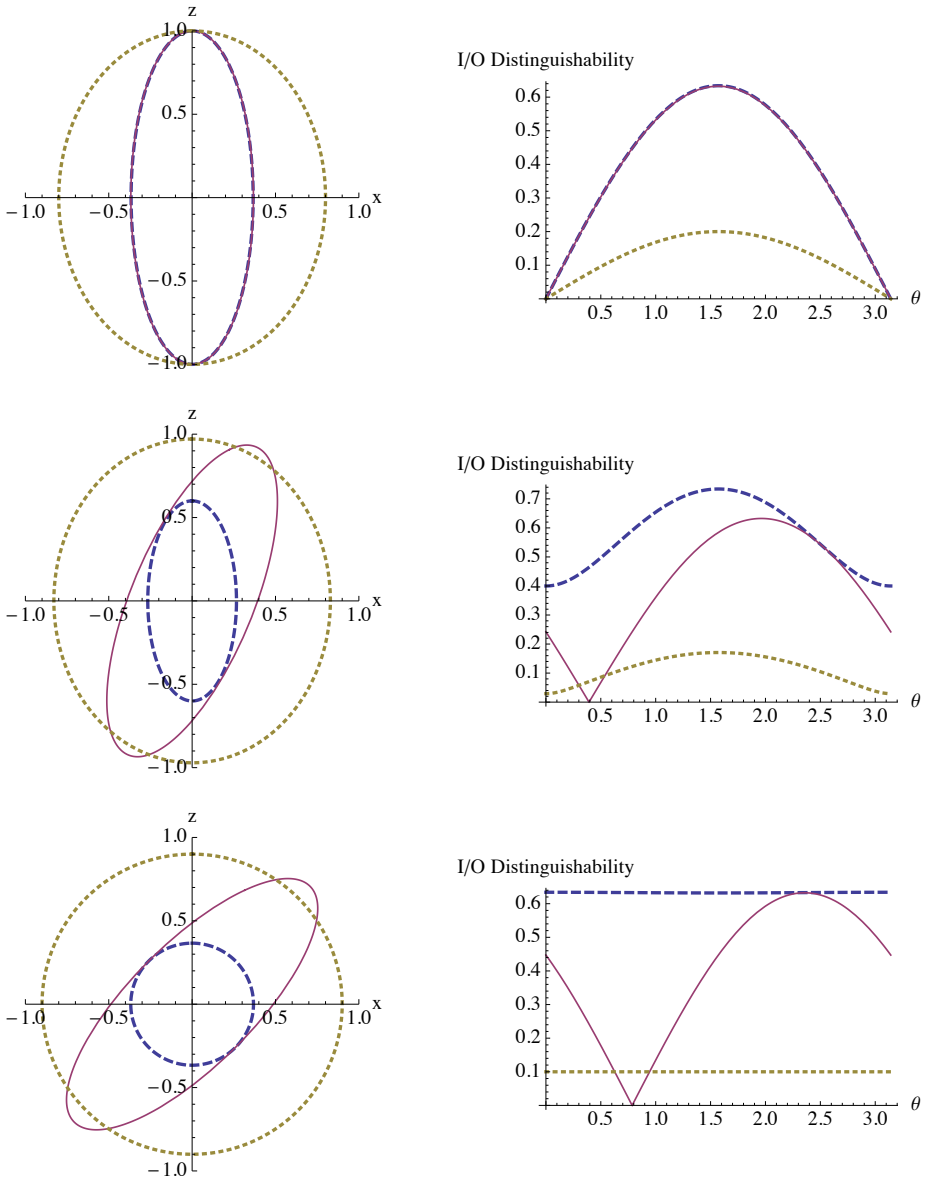


Figure 4.1: Bloch visualizations for the unitary examples and their Pauli and Pauli twirled approximations. From top to bottom, the channel (and its approximations) being represented are Λ_0, Λ_1, Λ_2. The solid red line represents the action of the Original channel, the dashed blue represents the Pauli, and the dotted gold represents the Pauli Twirled. The left column is a visualization of how the channels affect the x-z plane of the Bloch sphere, and the right column is a curve of the input-output distinguishability for the states in the x-z plane an angle θ from the z-axis.

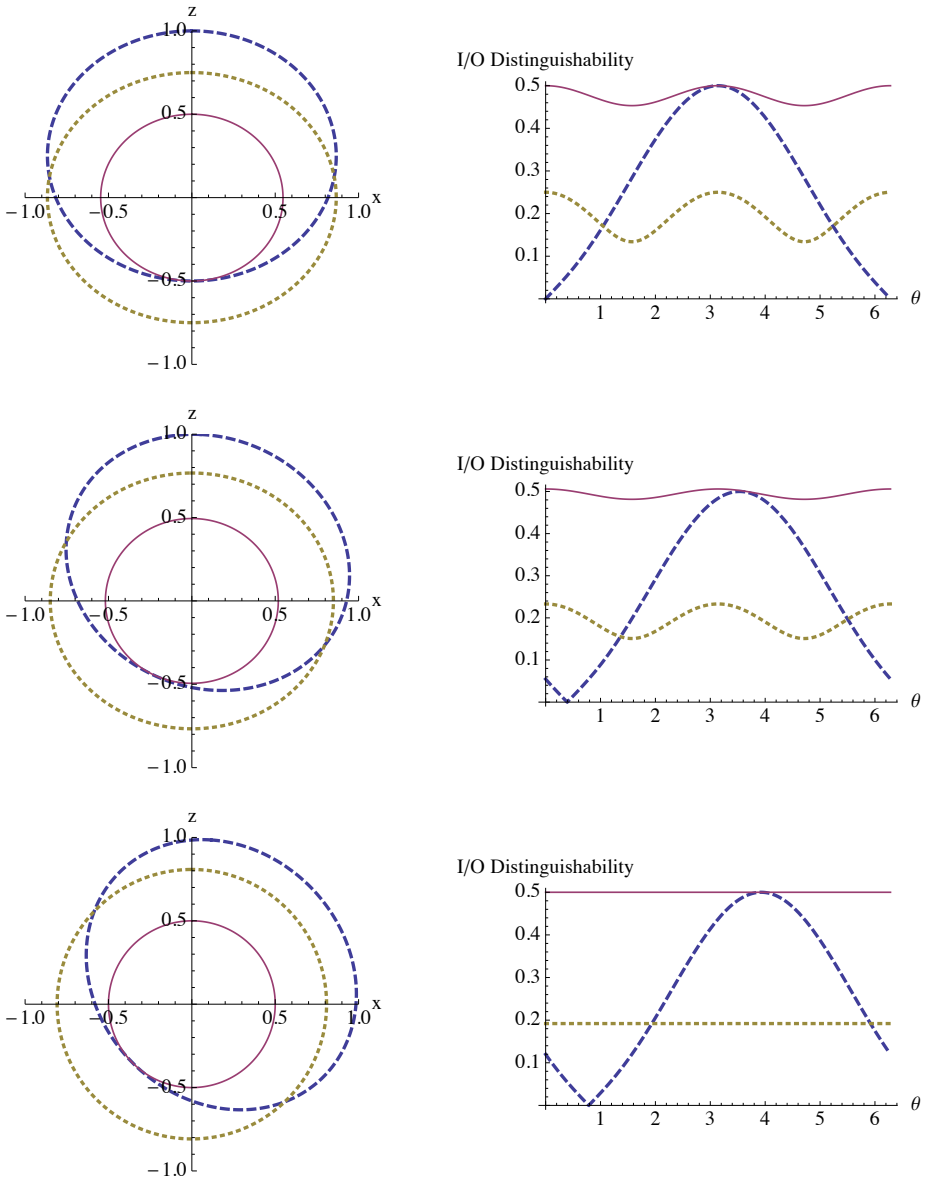


Figure 4.2: Bloch visualizations for the non-unital examples and their Pauli and Pauli twirled approximations. From top to bottom, the channel (and its approximations) being represented are $\mathcal{E}_0$, $\mathcal{E}_1$, $\mathcal{E}_2$. The solid red line represents the action of the Original channel, the dashed blue represents the Pauli, and the dotted gold represents the Pauli Twirled. The left column is a visualization of how the channels affect the x-z plane of the Bloch sphere, and the right column is a curve of the input-output distinguishability for the states in the x-z plane an angle θ from the z-axis.

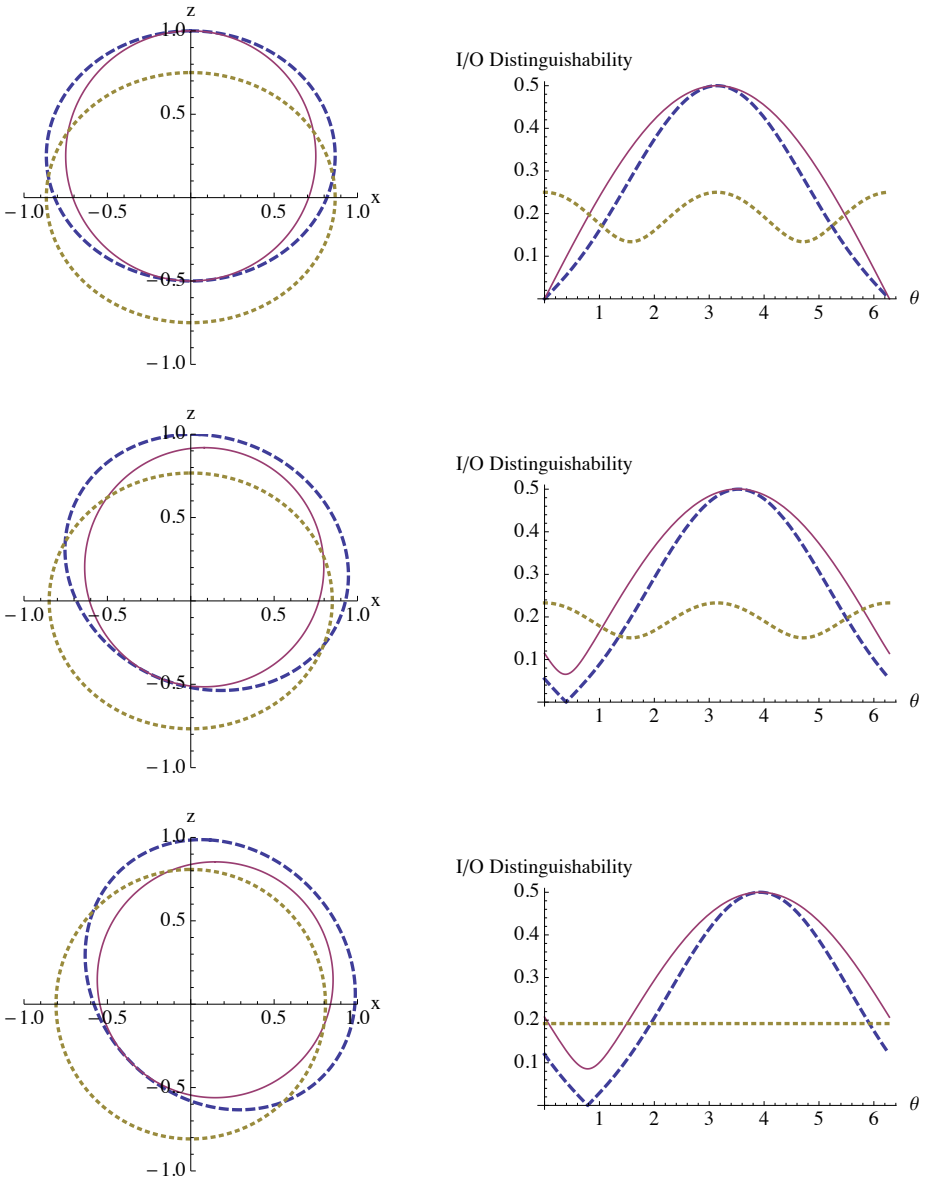


Figure 4.3: Bloch visualizations for the non-unital examples and their Pauli & Measurement and Pauli twirled approximations. From top to bottom, the channel (and its approximations) being represented are $\mathcal{E}_0$, $\mathcal{E}_1$, $\mathcal{E}_2$. The solid red line represents the action of the Original channel, the dashed blue represents the Pauli, and the dotted gold represents the Pauli Twirled. The left column is a visualization of how the channels affect the x-z plane of the Bloch sphere, and the right column is a curve of the input-output distinguishability for the states in the x-z plane an angle θ from the z-axis.

Chapter 5

Numerical Studies of Circuit Performance

Up until now, we've described and studied honesty of approximations in terms of the performance of individual gates. Ultimately however, we're interested in how the honest gate approximations can be used to understand circuit performance. We phrase this as the problem of *composition*. That is, while the individual gates may be honest, when they are put into a circuit, the "approximate" circuit may not be an honest representation of the original circuit.

In Section 5.1, we discuss the problem of composition in more detail and give a simple example that shows that honest approximations will not "honestly compose" in general. Nevertheless, despite the possible issues with generic error composition, it is possible that when considering realistic error models within typical QEC circuits, our approximations will compose honestly. In Section 5.2 and onward, we put our approximation method to the test within a work flow that strongly motivated this research. That is, within the context of understanding the performance of particular QEC circuits implemented on particular physical machines, as described in the introduction.

This chapter contains the main body of [12], slightly reorganized and modified for the flow of this thesis.

5.1 The Problem of Composition

We begin by constructing a simple example in which honest approximations compose dishonestly. This example demonstrates the inherent limitations of approximating an error with another error that composes in a fundamentally different way. Consider the two single-qubit maps

$$\Gamma(\rho) = U(\theta)\rho U^\dagger(\theta) \tag{5.1}$$
$$\Lambda(\rho) = (1-p)\rho + pZ\rho Z, \tag{5.2}$$

where $U(\theta) = \exp(-i\frac{\theta}{2}Z)$ and Z is the Pauli z operator. In [11], it was shown that if $p = |\sin(\theta/2)|$, then Λ and Γ have identical IO distinguishability properties; that is,

$$\|\rho - U(\theta)\rho U(\theta)^\dagger\|_1 = \|\rho - \Lambda(\rho)\|_1, \tag{5.3}$$

for all ρ, and therefore Λ is an honest representation of Γ. Consider the state $\rho_+ = |+\rangle\langle+|$, where $X|+\rangle = |+\rangle$, and X is the Pauli x operator, which is chosen as it is maximally sensitive to both $U(\theta)$ and Λ. One can check that

$$\|\rho_+ - \Gamma(\rho_+)\|_1 = \|\rho_+ - \Lambda(\rho_+)\|_1 = 2|\sin(\theta/2)| = 2p. \tag{5.4}$$

Now, consider the circuit composed of two applications of Γ, $C = \Gamma \circ \Gamma$, and the approximate circuit $C^{(a)} = \Lambda \circ \Lambda$, where

$$C^{(a)}(\rho) = [(1-p)^2 + p^2]\rho + 2p(1-p)Z\rho Z. \tag{5.5}$$

For $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, it can be checked that

$$\|\rho_+ - C(\rho_+)\|_1 > \|\rho_+ - C^{(a)}(\rho_+)\|_1, \tag{5.6}$$

and thus the composition of approximations is not an honest representation of the original circuit.

This example can be understood by looking at how repetitive application of these channels affects the state. Each application of Γ, a unitary error, deterministically rotates the state by a small angle, whereas Λ, a dephasing error, rotates the state by 180°, but with a small probability. In the former case, the distance that each application moves the current state remains constant, whereas in the latter case, this distance decreases exponentially in the number of applications, resulting in an underestimation of errors after only two applications. We encounter this situation in our simulations; one of the

gate sets we consider has an identity gate error that is essentially a unitary about the z-axis, and the honest Pauli approximation is the dephasing channel that reproduces its IO distinguishability properties. Given this discussion and the frequency with which the gate occurs, we expected that our approximations might underestimate the overall circuit error. However, even in this case, our approximations perform as desired, providing strong numerical evidence for the value of this method in QEC circuits.[1]

Despite these potential difficulties with error composition, it remains possible that, after a QEC protocol is applied, the resulting effective errors might compose more desirably. As an example, the first step of QEC in stabilizer codes is measuring the *error syndrome*. This consists of measuring a generating set of stabilizer elements $\{Q_i\}_{i=1}^k$, which produces a k-bit string b with $b_i = 1$ if the outcome from measuring Q_i is -1, and $b_i = 0$ if it is $+1$. If a particular string b is measured, then the system is projected onto the subspace defined by the projector

$$\Pi_b = 2^{-k}(\mathbb{1} + (-1)^{b_1}Q_1)\cdots(\mathbb{1} + (-1)^{b_k}Q_k). \tag{5.7}$$

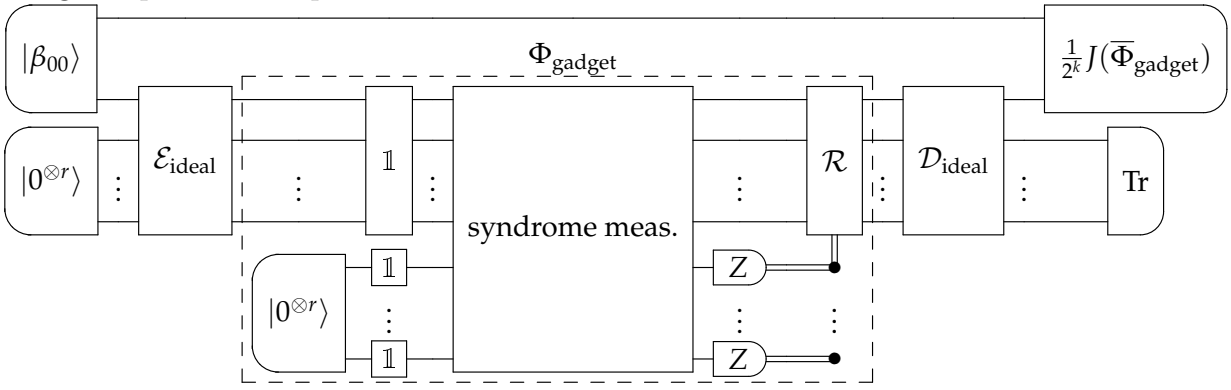
For a Pauli operator P and codeword $|\psi\rangle$, $\Pi_b P|\psi\rangle = 0$ if P does not produce syndrome b, and $\Pi_b P|\psi\rangle = P|\psi\rangle$ if P produces syndrome b. Thus, indexing the Pauli operators as $\{P_i\}$, and denoting S_b as the set of indices for Pauli operators that produce syndrome b, if a particular syndrome b is measured after an error Λ, with χ-matrix χ_{ij} in the Pauli basis [16], acts on an arbitrary codeword $|\psi\rangle$, the state will be (ignoring normalization)

$$\begin{aligned}\Pi_b\Lambda(|\psi\rangle\langle\psi|)\Pi_b &= \sum_{ij}\chi_{ij}\Pi_b P_i|\psi\rangle\langle\psi|P_j\Pi_b \\ &= \sum_{ij\in S_b}\chi_{ij}P_i|\psi\rangle\langle\psi|P_j.\end{aligned}$$

Given this form, it is clear that if P_i and P_j have different syndromes, then χ_{ij} can play no part in the post-syndrome measurement state, and is therefore effectively truncated by syndrome measurement. In this way, errors become more "incoherent" and this, at least superficially, makes errors "more like" Pauli channels (which have diagonal χ-matrices in the Pauli basis). Thus, whatever the form of Λ, after a syndrome extraction step is enacted,

[1] This example is for serial composition, but an interested example where honest parallel composition fails can be constructed in the following way. Take two unitary qubit maps, Λ_1, Λ_2, with the property that if they are composed in parallel on a maximally entangled state, the state is invariant. As is discussed in Section 4.4.1, there exist dephasing channels $\mathcal{E}_1$ and $\mathcal{E}_2$ that perfectly match the input-output distinguishability function of each unitary. Thus, individually, Λ_1 is a 1-honest representation of $\mathcal{E}_1$, and similarly for Λ_2 and $\mathcal{E}_2$. However, when composed in parallel on the maximally entangled state, labelled ρ, $0 = \|\rho - (\Lambda_1\otimes\Lambda_2)(\rho)\|_1 < \|\rho - (\mathcal{E}_1\otimes\mathcal{E}_2)(\rho)\|_1 \neq 0$.

Figure 5.1: Circuit to produce a Choi state for the logical action $\overline{\Phi}_{\text{gadget}}$ of a QEC gadget acting on an $[\![n, k, d]\!]$ stabilizer code, where $r \equiv n-k$. Note that errors act on all operations in the dashed box (including measurements), aside from the recovery operation labelled with $\mathcal{R}$. A detailed account of this circuit (including an explanation as to why $\mathcal{R}$ is left noiseless) is given in Section 5.5. Note that, while the Z measurements are part of syndrome measurement, we use the the box labelled "syndrome meas." to refer only to the gate operations required before actual measurement occurs.



the effective error may compose more like a Pauli channel. In practice, this argument may fail due to various aspects of imperfect syndrome measurement, such as limited visibility measurements, and the time it takes to perform measurement protocols like ancilla-assisted syndrome measurement.

5.2 Simulation Schema

In this section, we work through three examples of the workflow described in the introduction. Starting with two physical models (given in terms of deterministic and stochastic continuous-time evolution), we generate three sets of gates via various pulse design techniques and physical model parameters. Given these gate sets, we approximate the errors on each gate using our approximation method to create efficiently simulable honest gate sets. We then simulate a small, typical, QEC gadget, and compare how our approximate gate sets compare to the original, un-approximated gate sets. We find that in all cases, our approximations perform exactly as desired. We also include the performance of the Pauli twirled approximations.

Our simulation schema begins with a low level physical model, and ends with a high

level, efficient simulation of a QEC circuit, with our approximation method being a bridge between the two. A physical model is a description of the continuous-time dynamics of a candidate physical system for QIP, consisting of deterministic and stochastic parts to internal and control Hamiltonians, as well as dissipative open quantum-system dynamics, and includes constraints on control amplitudes. Using control techniques, a gate set is generated from a physical model, which consists of all elementary quantum logic gates required for the desired QEC circuit. Once a gate set is generated, the error on each gate is approximated using our method, yielding an *honest* representation of the original gate set, which is used in an efficient simulation of the desired QEC circuit.

The circuit that we simulate, shown in Figure 5.1, is a gadget that performs one round of error-correction in an $[\![n, k, d]\!]$ stabilizer code, which we implement using the 5-qubit perfect code (a $[\![5, 1, 3]\!]$ code). The circuit was chosen by balancing the desire that it be representative of standard practices, with the requirement that it be small enough to allow for fast simulation of arbitrary gate errors, thereby allowing for comparison of the efficiently simulable errors to the original. For us, a gate set consists of the gates $\{\mathbb{1}, X, Y, Z, H, \text{CNOT}\}$, where the first four are standard single qubit Pauli gates, H is the Hadamard gate, and CNOT is the two qubit controlled-not gate.[2]

We implement this procedure for three gate sets, generated from two physical models with varying model parameters and control techniques. The physical models and control techniques are chosen to be representative of those found in experiment. Doing so allows us to encounter errors not typically considered in fault-tolerance research, despite naturally occurring in physical implementations. We emphasize however that neither the approximation method, nor the procedure for testing it given here, have been tailored for a particular outcome. The method is generic; it is independent of both the underlying physical model and gates, as well as the QEC circuit.

5.3 Gate Set Generation

We consider two physical models, PM1 and PM2. PM1 is motivated by a double quantum dot system, and PM2 represents an archetypal two level system (see Appendix E for a description and full details). Gate Set 1 (GS1) is built on PM1, and Gate Sets 2 and 3 (GS2 and GS3) are built on PM2 (using different model parameters). GS1 and GS2 use noise refocusing techniques [30, 31], which mitigate errors induced by stochastic Hamiltonians.

[2]Note that the circuit we ultimately simulate uses only the $\mathbb{1}$, H, and CNOT gates. The rest are included for further comparison of the gate approximations.

GS2 and GS3 implement gates via hard pulses; that is, the pulse sequences used to generate the gates are manually specified by choosing control amplitudes. Due to the complicated structure of the Hamiltonian in PM1, optimal control theory (OCT) was used to find pulse sequences that implement the gates in GS1 with high fidelity [32, 33]. Every gate in a set is made to be the same length in time, as our circuit simulation proceeds in discrete time steps in which a single gate acts on every register qubit. A full description of how the gates are simulated is given in Appendix F.1.

The different combinations of physical model and control techniques give rise to different types of gate errors. A detailed account on the form of the errors for each gate set is given in Chapter 6, as it has particular relevance within the context of that discussion. We do however wish to highlight that some gates in GS1 have largely unitary errors, resulting from the use of OCT pulse finding in gate implementation. Thus, given the discussion on error composition, GS1 provides a strong test for our method.

5.4 Gate Set Approximations and Statistics

For each gate set we generate three efficiently simulable approximate gate sets:

Pauli twirled The Pauli twirled errors.

Pauli The honest Pauli channel approximation.

Clifford The honest mixed-Clifford channel approximation.[3]

Table 5.1 presents the statistics for the identity gate from each gate set, using $N = 10^6$ uniformly sampled pure states (see Section 4.3 for a description of the displayed quantities). See Appendix D for tables containing statistics on all gates.

Looking at the various diamond norm distances between the Original, Ideal, Pauli twirled (Λ_{PT}), and Pauli errors (Λ_{P}), a simple ordering can be seen to hold for every gate:

$$\|\Lambda_{\mathrm{PT}} - U_{\mathrm{Ideal}}\|_\diamond \leq \|\Lambda_{\mathrm{Original}} - U_{\mathrm{Ideal}}\|_\diamond \leq \|\Lambda_{\mathrm{P}} - U_{\mathrm{Ideal}}\|_\diamond ,$$
$$\|\Lambda_{\mathrm{PT}} - \Lambda_{\mathrm{Original}}\|_\diamond \leq \|\Lambda_{\mathrm{P}} - \Lambda_{\mathrm{Original}}\|_\diamond .$$

[3]Note that only the single-qubit errors are approximated as mixed-Cliffords (we allowed the algorithm to search over all mixed-Cliffords). Due to the large number of two-qubit Cliffords, the honest Pauli approximation for the CNOT gate is reused.

Table 5.1: Statistics for the various approximations of the identity gate for each gate set, approximated from $N = 10^6$ random pure states.

Set/Statistics		Original	Pauli twirled	Pauli	Clifford
GS1	χ_{00}	0.999994	0.999994	0.997618	0.998314
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.76×10^{-3}	1.20×10^{-5}	4.76×10^{-3}	4.77×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.76×10^{-3}	6.73×10^{-3}	3.64×10^{-3}
	$\overline{h}$		-3.73×10^{-3}	1.14×10^{-7}	1.64×10^{-6}
	p_{viol}		1.	0.	0.
GS2	χ_{00}	0.999087	0.999087	0.999085	0.999086
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.48×10^{-5}	2.50×10^{-5}	2.06×10^{-6}
	$\overline{h}$		-1.63×10^{-7}	2.65×10^{-6}	8.34×10^{-7}
	p_{viol}		0.49861	0.	0.
GS3	χ_{00}	0.998751	0.998751	0.996501	0.996501
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.99×10^{-3}	2.50×10^{-3}	7.00×10^{-3}	7.00×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.50×10^{-3}	5.28×10^{-3}	5.28×10^{-3}
	$\overline{h}$		-1.03×10^{-3}	1.91×10^{-3}	1.91×10^{-3}
	p_{viol}		0.74978	0.	0.

Thus, while the Pauli twirled error is always closer than the Pauli to the Original, it is also always closer to the Ideal than the Original, and is therefore a less noticeable error than the Original. The location of the Clifford approximation in the second inequality chain varies; for some gates, it is an order of magnitude closer to the Original than both the Pauli and Pauli twirled approximations, for others, it is the same distance to the Original as the Pauli, indicating that the best Pauli is also the best Clifford approximation, and for the rest, it is in between the two.

Two other important and connected observations can be made. In some cases, the χ_{00} element of the honest approximations is much lower than that of the Original, and in others it is not appreciably different. In the former cases, the Pauli twirled approximations tend to be much closer to the Ideal, and have worse hedging performance, than in the latter cases. These are demonstrations of channels with different average fidelities but similar IO distinguishability properties, and channels with identical average fidelities but very different IO distinguishability properties. This observation is connected to the different regimes of performance for Pauli twirling, as well as how "coherent" or "unitary" an error is, and is explained in detail in Chapter 6.

5.5 Circuit Design and Simulation Results

We simulate the gadget Φ_{gadget} that performs one round of error correction on one block of an error-correcting code, as per Figure 5.1. We isolate the action $\overline{\Phi}_{\text{gadget}}$ of this gadget on the encoded state by preceding and following it with perfect encoding and decoding operations, $\mathcal{E}_{\text{ideal}}$ and $\mathcal{D}_{\text{ideal}}$[4]. The circuit is simulated by computing its Choi state; for a code that encodes k logical qubits into n physical qubits, we take the state

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2^k}} \sum_{i=1}^{2^k} |i\rangle \otimes |i\rangle \,, \tag{5.8}$$

where $\{|i\rangle\}_{i=1}^{2^k}$ is an orthonormal basis for the k-qubit Hilbert space, and compute

$$\left(C \otimes \mathbb{1}_{L(\mathbb{C}^k)}\right) \left(|\beta_{00}\rangle \langle\beta_{00}|\right) = \frac{1}{2^k} J(C), \tag{5.9}$$

where C represents the entire circuit and $J(C)$ is the Choi-Jamiołkowski matrix for the circuit. (This simulation method is chosen due to the efficiency with which maps of the

[4]Note that sandwiching gadgets in a circuit between perfect encoding and decoding operations will not guarantee that the approximate gadgets will compose honestly.

form we consider can be applied to states. See Appendix F.2 for details.) Due to the perfect encoding and decoding operations, $J(C) = J(\overline{\Phi}_{\text{gadget}})$. Explicitly, the circuit performs the following operations:

1. The state $|\beta_{00}\rangle$ (Equation 5.8) is prepared, half of which is perfectly encoded into an $[\![n, k, d]\!]$ stabilizer code.

2. The gadget Φ_{gadget} is applied to the encoded physical qubits, consisting of:

   i) One imperfect wait location on all of the data qubits. This is a placeholder for possible non-trivial operations in gadgets meant to perform logical operations.

   ii) Simultaneously, a register of ancillas for ancilla-assisted syndrome measurement is prepared. An imperfect identity operation acts on each ancilla to represent imperfect ancilla preparation.

   iii) Imperfect ancilla-assisted syndrome measurement is performed (see Figure 5.3). Measurement of the physical ancillas is taken to be perfect, with errors represented by identity gates that precede the measurement.

3. A perfect recovery operation is performed by classical feed-forward of syndrome measurement details (see Figure 5.4).

4. Once Φ_{gadget} is done, the resultant state is then perfectly decoded and the physical ancillas are discarded.

The recovery operation is chosen to be perfect as, in practice, it isn't always necessary to physically perform the recovery; errors can be tracked and taken into account when further operations are performed on the block [34].

We implement this simulation schema using the 5-qubit perfect code, with the syndrome measurement and recovery circuits designed by the Python package QuaEC [35]. Note that we choose to perform the syndrome measurement in a non-fault-tolerant way, as a fault-tolerant gadget for a code with n physical qubits would require at least $2n$ ancilla qubits for the Steane or Knill fault-tolerant error correction (FTEC) gadgets, or strictly more than $\sum_i \text{wt}(S_i) = 16$ ancillae for the Shor FTEC gadget. Thus, at least 10 ancillae are needed for the perfect code, requiring simulation of at least 16 qubits, putting us outside the range of quickly simulable circuits with arbitrary errors.

With 4 stabilizer generators, this code requires 4 ancillas for encoding and 4 for syndrome measurement using the circuit shown in Figure 5.3. Any redundant Hadamard gates have been removed. The recovery operation is shown in Figure 5.4. This type of non-fault

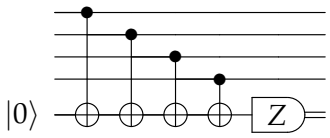


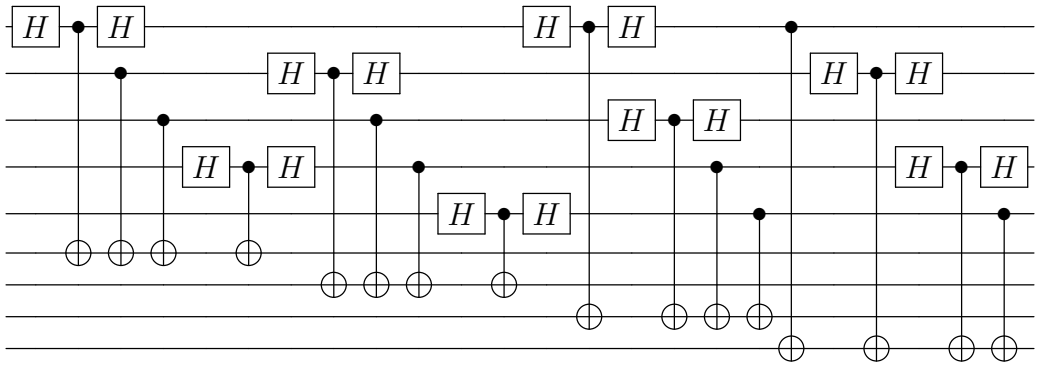
Figure 5.2: Circuit to measure the stabilizer generator $Z^{\otimes 4}$, proposed by Fowler et al [1] for use in surface codes.

tolerant syndrome measurement is similar to gadgets proposed for use in topological QEC codes, such as the surface code. In particular, the syndrome measurement gadget used by Fowler et al [1], shown in Figure 5.2, relies on CNOT gates between each data qubit in the support of a stabilizer generator and a common ancilla qubit. We emphasize that we are not concerned with *absolute* circuit performance. Rather, the task at hand is the comparison of *relative* performance of efficiently simulable error approximations, so it suffices that this circuit contains all of the typical elements and procedures for QEC, regardless of fault-tolerance.

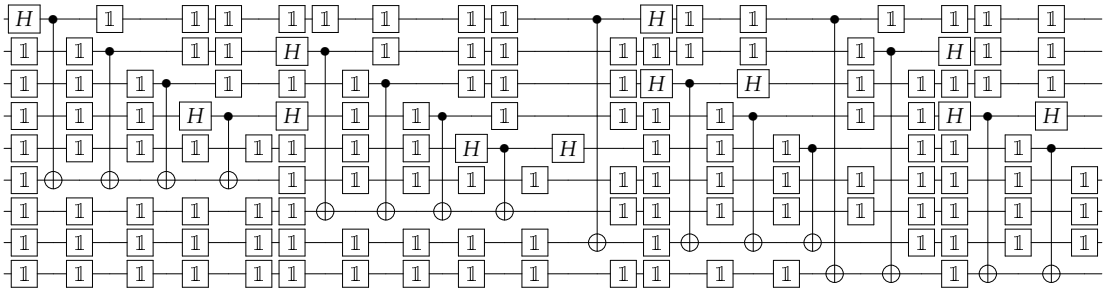
Tables 5.2 and 5.3 give the simulation statistics for $\overline{\Phi}_{\text{gadget}}$, again using $N = 10^6$ sample pure states to compute the hedging parameters. For each gate set, the Pauli and Clifford approximations compose well; the approximated circuit honestly represents the error of the original. This is especially encouraging for GS1, given its unitary identity error. For the Pauli twirled errors, we see that in GS1 they fail the honesty condition for every tested pure state. For GS2, they fail the honesty condition for just over half of the pure states tested, but by an arguably small degree. Interestingly, for GS3, the Pauli twirled approximations provide an honest representation of the circuit performance.

5.6 Discussion

In all examined cases, the honest approximations led to honest representations of circuit performance, providing confidence in our method as a tool for evaluating the performance of typical QEC circuits with realistic gate errors. By starting from continuous-time physical models, and building gates using common control techniques, we tested our method against errors typical of those found in experiment, which were not tailor-made for any desired outcome. The strongest test of our method came from errors with strong unitary parts, arising from OCT designed pulses, a regime not typically considered in fault-tolerance research. Additionally, our results, in conjunction with the recent work by Geller and Zhou [36], demonstrate two regimes of performance for Pauli twirled error approximations.



(a) Original syndrome measurement circuit.



(b) Circuit with simplifications and with explicit wait locations.

Figure 5.3: Syndrome measurement circuit for the five-qubit perfect code.

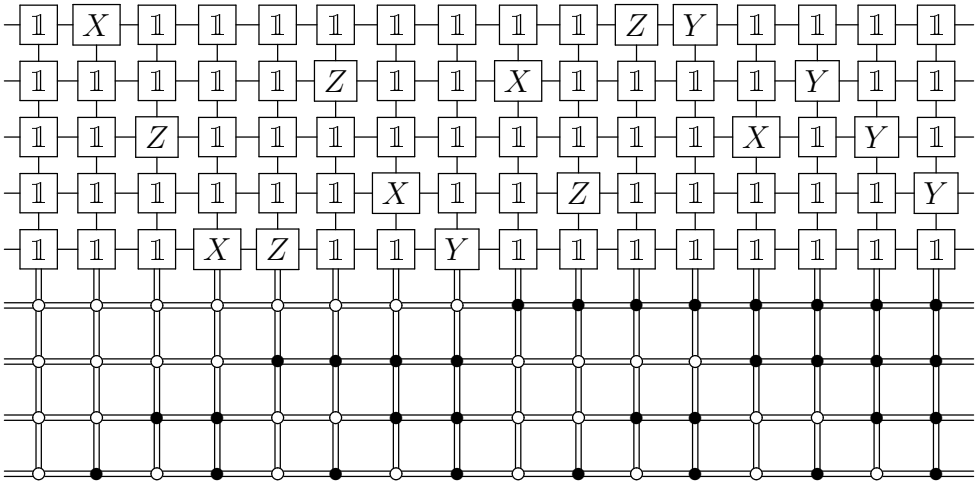


Figure 5.4: Recovery circuit for the five-qubit perfect code.

Table 5.2: Statistics for $\overline{\Phi}_{\text{gadget}}$ using $N = 10^6$ randomly sampled pure states. The key point to notice is that $p_{\text{viol}} = 0$ for all approximations that were generated using our method. That is, the action of the circuit on all sampled pure states was found to be 1-honest. This is summarized in Table 5.3.

Set/Statistics		Original	Pauli twirled	Pauli	Clifford
GS1	χ_{00}	0.999964	0.999964	0.985820	0.989930
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.76×10^{-3}	7.28×10^{-5}	2.84×10^{-2}	2.04×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.76×10^{-3}	2.87×10^{-2}	2.01×10^{-2}
	$\overline{h}$		-3.69×10^{-3}	1.85×10^{-2}	1.23×10^{-2}
	p_{viol}		1.	0.	0.
GS2	χ_{00}	0.991372	0.991372	0.991355	0.991367
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.73×10^{-2}	1.73×10^{-2}	1.73×10^{-2}	1.73×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.45×10^{-5}	4.29×10^{-5}	1.14×10^{-5}
	$\overline{h}$		-1.63×10^{-8}	2.24×10^{-5}	7.66×10^{-6}
	p_{viol}		0.55566	0.	0.
GS3	χ_{00}	0.992495	0.987594	0.969499	0.969499
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.51×10^{-2}	2.48×10^{-2}	6.10×10^{-2}	6.10×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.03×10^{-2}	4.60×10^{-2}	4.60×10^{-2}
	$\overline{h}$		6.36×10^{-3}	3.04×10^{-2}	3.04×10^{-2}
	p_{viol}		0.	0.	0.

Table 5.3: A summary of Table 5.2. If the approximation performed as desired (produced a 1-honest representation of the original circuit) then it gets a 3, and if not, an 7. For the Pauli twirled approximation for GS2, we put both a 3 and an 7, as while approximately half of the sampled states failed the honesty condition, the underestimation was small.

Set	Pauli twirled	Pauli	Clifford
GS1	✗	✓	✓
GS2	✓/✗	✓	✓
GS3	✓	✓	✓

In one regime, their performance can be considered "sufficiently good", while in the other, Pauli twirling results in systematic underestimation of the IO distinguishability notion of error (see Chapter 6 for a systematic investigation of this observation).

Our work is motivated by the desire for the simulations to be pessimistic. We want to be reasonably assured that, if the simulation with the approximated errors performs well according to some metric, then the actual implementation will perform well also. Currently, experimental implementations of QIP are limited to small system sizes. Extrapolating their performance to hypothetical large-scale systems requires caution. Quantum processors will require constant application of error-correction protocols like the one we consider, and it is imaginable that in large systems, consisting of hundreds of qubits or more, even a small underestimation of the effect of physical-level errors may dramatically compound, resulting in false expectations of over-all performance. This work provides hope that our method can be used as a tool for extrapolating performance from small systems in an honest way.

Chapter 6

Analysis of the Form of Errors

This chapter contains an analysis of how the form of the original errors affects the various approximations that we have considered throughout this thesis. In particular, we identify good and poor regimes of performance for the "Pauli twirled" errors, and classify the gate sets of the preceding chapter into these regimes. This also aids in the understanding of the behaviour of the approximations generated by our own method; in particular the observations made in Sections 4.4 and 5.4 that, in some cases, the average fidelity of our approximations is much less than that of the original, whereas in some cases they are very similar. Given the importance on the classification of errors to this discussion, we conclude by explaining why the errors for each gate set in the preceding chapter take the form that they do. We note that this chapter is Appendix C in [12], with minor modifications for incorporation into this thesis.

The identification of these regimes requires analysis on what happens to the IO distinguishability properties of a channel when it is twirled. Generally, twirling a map Λ by a set of unitaries $\{U_k\}_{k=1}^N$ is the action of mapping $\Lambda \to \frac{1}{N}\sum_{k=1}^N U_k^\dagger \circ \Lambda \circ U_k$. The twirled map results from choosing a unitary operator from the twirling set with uniform probability, applying it, applying Λ, then inverting the twirling operator. If the twirling set is chosen to be the Pauli operators, it is called a Pauli twirl and, if perfectly implemented, will transform any map into the Pauli channel that results from mathematically truncating the off-diagonal elements of the process (χ-)matrix [16].

Before analyzing the effects of Pauli twirling specifically, we can look at the general effect of twirling on the diamond norm distance of an arbitrary channel to the identity operation. Let $\mathcal{H}$ denote a finite dimensional Hilbert space, and $L(\mathcal{H})$ the set of linear operators from $\mathcal{H} \to \mathcal{H}$. One property of the diamond norm is that for $\Phi : L(\mathcal{H}) \to L(\mathcal{H})$,

and any unitary operators $U, V \in L(\mathcal{H})$, it holds that $\|U \circ \Phi \circ V\|_\diamond = \|\Phi\|_\diamond$ [13]. Thus, for any finite set of unitaries $\{U_k\}_{k=1}^N \subset L(\mathcal{H})$, and quantum channel $\Lambda : L(\mathcal{H}) \to L(\mathcal{H})$, it holds by straightforward application of the triangle inequality that

$$\left\| \mathbb{1}_{L(\mathcal{H})} - \frac{1}{N} \sum_{k=1}^{N} U_k^\dagger \circ \Lambda \circ U_k \right\|_\diamond \leq \left\| \mathbb{1}_{L(\mathcal{H})} - \Lambda \right\|_\diamond , \tag{6.1}$$

where $\mathbb{1}_{L(\mathcal{H})}$ is the identity channel. In other words, the distance of a twirled error to the identity operation is always bounded above by that of the original error and so, in a worst case sense, twirling typically acts to make an error harder to detect.

To specify the regimes of performance of twirling, we look at the Bloch representation of quantum channels. Any qubit map Λ can be represented as a matrix M and vector $\vec{t}$ that acts on Bloch vectors as $\vec{r} \to M\vec{r} + \vec{t}$. M can be written in terms of its polar decomposition $M = OP$, where O and P are an orthogonal and positive semidefinite matrix, respectively. Thus, the action on the Bloch sphere can be represented as $\vec{r} \to O(P\vec{r} + O^T\vec{t})$ [37]. That is, as a possibly non-unital channel followed by an orthogonal rotation of the Bloch sphere, which corresponds to a unitary rotation for qubits [38]. We say that a channel is in the "unitary regime" if the effect of O is relatively large compared to P and $\vec{t}$, and we say the channel is in the "deforming regime" if the opposite is true. To quantify the "effect" of a Bloch matrix M, we use the quantity $\|\mathbb{1} - M\|_2$, and use $\|\vec{t}\|_2$ to quantify the effect of the non-unital part. We use $\|\cdot\|_2$ to denote both the Hilbert-Schmidt norm on matrices and the Euclidean norm on vectors, where context and notation will make clear which is meant. In the following paragraphs, we examine how twirling affects worst-case IO distinguishability properties of different errors in these regimes.

First, for a unital deforming qubit channel Λ, the Bloch representation is simply a positive semi-definite matrix P. Diagonalize P as $P = UDU^\dagger$, for some orthogonal matrix U and diagonal non-negative matrix D. Then, as diagonal Bloch matrices are realizable as Pauli channels, we can write Λ as $\Lambda(\rho) = \sum_{i=0}^3 p_i V P_i V^\dagger \rho V P_i V^\dagger$, where V is the unitary corresponding to U. Channels of this form are called Generalized Pauli channels, and the diamond norm distance between channels of this form with different probability vectors $\vec{p}$ and $\vec{q}$ is given by $\sum_{i=0}^3 |p_i - q_i|$ [39, 40]. For these channels, as $VP_iV^\dagger$ has no identity part for $i \geq 1$, the χ_{00} element in the Pauli basis is identical to the probability assigned to the identity operator. As this quantity is conserved in Pauli twirling, and as the identity map on qubits $\mathbb{1}_{L(\mathbb{C}^2)}$ is a Generalized Pauli channel, it follows that, for a deforming unital qubit channel Λ and its Pauli twirl Λ_{PT}, $\|\mathbb{1}_{L(\mathbb{C}^2)} - \Lambda\|_\diamond = (1 - p_0) + p_1 + p_2 + p_3 = 2(1 - p_0) = \|\mathbb{1}_{L(\mathbb{C}^2)} - \Lambda_{\mathrm{PT}}\|_\diamond$. Thus, diamond norm distance to the identity for these channels is unaffected by Pauli twirling.

To examine the qualitative behaviour of Pauli twirling on a non-unital deforming channel, we examine the special case of an amplitude damping channel, which has Kraus operators

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \; K_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \tag{6.2}$$

for some parameter γ, and Bloch representation

$$M = \begin{pmatrix} \sqrt{1-\gamma} & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 \\ 0 & 0 & 1-\gamma \end{pmatrix}, \; \vec{t} = \begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix}. \tag{6.3}$$

In the Bloch representation, the effect of Pauli twirling is to remove the off-diagonal elements of M and set $\vec{t} \to 0$. Thus, the only effect that twirling has on this channel is to remove the non-unital part. Note that the dominant error considered by Geller and Zhou in [36] is of this form; the Bloch matrix is diagonal, and so in some sense, Pauli twirling has a minimal effect. We consider the worst case performance of this channel on a pure qubit state. It is clear that the state most affected by Λ, and by its Pauli twirl Λ_{PT}, is the -1 eigenstate of Z, which we denote as ρ_-. It can be easily checked that $\|\rho_- - \Lambda(\rho_-)\|_1 = 2\gamma$, and $\|\rho_- - \Lambda_{\mathrm{PT}}(\rho_-)\|_1 = \gamma$. Thus, while the worst case performance in this case is lessened by Pauli twirling, it is only by a factor of 2.

Lastly, we look at purely unitary channels. For two unitary operators U, V in $L(\mathcal{H})$, there exists a pure state $|\psi\rangle \in \mathcal{H}$ for which

$$\|U \cdot U^\dagger - V \cdot V^\dagger\|_\diamond = \|U |\psi\rangle \langle\psi| U^\dagger - V |\psi\rangle \langle\psi| V^\dagger\|_1 = 2\sqrt{1 - |\langle\psi| U^\dagger V |\psi\rangle|^2} \tag{6.4}$$

[13]. From this form, it is clear that any state that maximizes the distinguishability between the identity operation and a unitary U will be orthogonal to the unitary's rotation axis. Thus, for a rotation U by an angle θ, $\|\mathbb{1}_{L(\mathbb{C}^2)} - U \cdot U^\dagger\|_\diamond = 2|\sin(\theta/2)|$. As the χ_{00} element of a qubit rotation by angle θ is $\cos^2(\theta/2)$, it follows from the preceding discussion that, for the Pauli twirled error Λ_{PT}, $\|\mathbb{1}_{L(\mathbb{C}^2)} - \Lambda_{\mathrm{PT}}\|_\diamond = 2(1 - \cos^2(\theta/2)) = 2\sin^2(\theta/2)$. Thus, for small values of θ, the twirled channel can be orders of magnitude closer to the identity than the original.

With this qualitative analysis in hand, we examine the form of the Bloch representation of the errors considered in this paper. The Bloch representation of each error is decomposed into the three pieces O, P, and $\vec{t}$, and the size of each piece is reported in Table 6.1. As the identity gate occurs most frequently in the circuit, its properties are the most important.

Table 6.1: We denote the Bloch representation of each gate error as $(OP, \vec{t})$, where O is an orthogonal matrix, P is positive-semidefinite, and $\vec{t}$ is the "non-unital part". Norms of size less than 10^{-10} are displayed as 0, as at this size they are irrelevant compared to the dominant parts of the error.

Set/Statistics		$\mathbb{1}$	X	Y	Z	H	CNOT
GS1	$\|\mathbb{1} - O\|_2$	6.74×10^{-3}	1.83×10^{-2}	1.83×10^{-2}	5.06×10^{-3}	1.63×10^{-2}	5.78×10^{-2}
	$\|\mathbb{1} - P\|_2$	8.22×10^{-7}	1.06×10^{-2}	4.43×10^{-3}	1.33×10^{-2}	6.24×10^{-3}	3.56×10^{-2}
	$\|\vec{t}\|_2$	0.	5.51×10^{-9}	5.56×10^{-9}	0.	1.36×10^{-8}	2.19×10^{-8}
GS2	$\|\mathbb{1} - O\|_2$	3.35×10^{-5}	5.52×10^{-4}	5.52×10^{-4}	2.09×10^{-4}	1.59×10^{-3}	1.73×10^{-3}
	$\|\mathbb{1} - P\|_2$	2.16×10^{-3}	2.15×10^{-3}	2.15×10^{-3}	2.32×10^{-3}	2.19×10^{-3}	8.13×10^{-3}
	$\|\vec{t}\|_2$	4.01×10^{-8}	5.65×10^{-5}	4.95×10^{-5}	1.66×10^{-4}	9.91×10^{-5}	2.14×10^{-4}
GS3	$\|\mathbb{1} - O\|_2$	0.	1.09×10^{-5}	1.50×10^{-5}	1.04×10^{-5}	2.83×10^{-6}	4.45×10^{-2}
	$\|\mathbb{1} - P\|_2$	3.06×10^{-3}	2.93×10^{-3}	2.93×10^{-3}	3.06×10^{-3}	2.96×10^{-3}	1.12×10^{-2}
	$\|\vec{t}\|_2$	2.50×10^{-3}	1.37×10^{-3}	1.37×10^{-3}	2.50×10^{-3}	2.11×10^{-3}	2.95×10^{-3}

We see that GS1 has the largest unitary component to its errors (in terms of the ratio to the other components). The error in the identity gate is almost entirely unitary, and therefore falls neatly into the unitary regime. Indeed, this is consistent with the fact that a single-qubit unitary rotation by an angle θ will have a χ_{00} element of $\cos^2(\theta/2) \approx 0.99999$, whereas the Pauli channel that *exactly* matches its IO distinguishably has a χ_{00} element of $1 - |\sin(\theta/2)| \approx 0.99762$. We see that these numbers correspond exactly to the given number of digits in Table 5.1, where a two-order-of-magnitude decrease in the distance of the Pauli twirled error to the identity channel is shown. While the other errors in the set have unitary parts, the non-unitary parts are comparable in size. As such, character from both the unitary and unital deforming regimes is observed; the diamond norm distance of the Pauli twirled error to the identity decreases, but the decrease is not as impressive as for the identity gate. For the Y and H errors, the unitary part is 2 or 3 times larger than the non-unitary, and thus the decrease in diamond norm distance to the ideal channel for the Pauli twirled errors is the greatest of the non-identity gates.

For every gate in GS2, $\|\vec{t}\|_2$ is relatively small and $\|\mathbb{1} - O\|_2$ is in most cases an order of magnitude smaller than $\|\mathbb{1} - P\|_2$, putting GS2 into the unital deforming regime. When looking at the diamond norm distances of the Pauli twirled channels to the identity, we see that, as expected in this regime, there is very little decrease. Indeed, while the Pauli twirled channels generally underestimate the error in the hedging metrics, the underestimation is small.

For GS3, $\|\mathbb{1} - O\|_2$ is usually of comparatively negligible size, and $\|\vec{t}\|_2$ is of the same order of magnitude as $\|\mathbb{1} - P\|_2$, putting this gate set into the non-unital deforming regime. Looking at the decrease of the diamond norm distance to the ideal operations for the Pauli twirled errors, we see roughly what is expected; a decrease on the order of a factor of 2. As in the case of GS2, while the Pauli twirled errors show underestimation in the hedging statistics, it is by a small degree.

The performance of the gadget, as simulated using the Pauli twirled approximations to the various gate sets, is more-or-less in-line with what might be expected from the preceding discussion of these regimes. For GS1, in the unitary regime, we observe large underestimation of IO distinguishability properties on an individual gate basis, and this underestimation propagates upwards; the gadget simulated with Pauli twirled errors is much harder to distinguish from the perfect gadget than the gadget subject to the Original errors. In the deformation regime (GS2 and GS3), where the Pauli twirled errors may only slightly underestimate the Original gate errors, we see very minor underestimation of gadget errors in GS2. One might argue, as Geller and Zhou do, that this level of underestimation is acceptable. Indeed, for GS3, the Pauli twirled circuit even has positive hedging.

These observations are perhaps not surprising on an intuitive level. Pauli channels are a subset of deforming channels, and while not all deforming errors are Pauli channels, they are *like* Pauli channels. All deforming channels contract the Bloch sphere inwards (with possible non-unital shifts), so Pauli twirling simply maps one deforming channel to another, which doesn't dramatically change the IO distinguishability properties. This stands in contrast to the unitary regime where, in regards to IO distinguishability, something fundamental can be lost when a dominantly unitary error is mapped to a Pauli channel via Pauli twirling.

In addition to understanding the performance of the Pauli twirled errors, this discussion also demonstrates why our approximations behave the way they do in regards to average gate fidelity. As we've seen from the Pauli twirl analysis, for a fixed average fidelity, channels in the unitary regime are much more distinguishable from the identity operation than those in the deforming regime. Thus, if the original error is in the unitary regime, our approximations, which fall into the deforming regime, will sometimes have a much worse average fidelity than the original to allow them to honestly represent its IO distinguishability properties.

As a final point, we can ask why the errors for each gate set have the form that they do. To explain this, it is necessary to examine the physical models and the pulse design techniques used, which are described in detail in Appendix E. The simplest case is GS3, where hard pulses with no refocusing sequences were used. As the T_1 time was the shortest

of all of the gate sets, and no refocusing pulses were used, the strong non-unital effects were allowed to continue building in their natural direction throughout each gate resulting in primarily non-unital errors. In GS2, which used the same physical model as GS3, the T_1 time was an order of magnitude longer. In addition, the refocusing pulse sequence used has the effect of "flipping" the non-unital shift between the positive and negative z-direction, and so, when averaged over the whole sequence, the effective non-unital shift tends to zero. As a result, the size of the non-unital piece tends to be at least an order of magnitude smaller than the other pieces, putting this gate set into the unital deformation regime. Finally, as GS1 had the longest T_1 time, and the refocusing sequence was performed at a rate far faster than the T_1 time, the non-unital error components are negligible. Aside from this, the prime difference between GS1 and the other gate sets is that the pulses used were found using the GRAPE algorithm, rather than being hard, which accounts for the dominant unitary components of the errors. For hard pulses, in the limit of no noise, the implemented unitary should be perfect, aside from numerical imprecision in the field amplitudes, which can be made arbitrarily small with low overhead. For pulses found using the GRAPE algorithm this is not the case; even a noiseless implementation will not produce the perfect unitary. Arguably, this situation is more prevalent in experimental implementations as pulse finding algorithms like GRAPE are necessary to obtain high-fidelity control over systems with complicated Hamiltonians. In addition, these types of algorithms can be used to achieve high-fidelity control across a range of internal Hamiltonian parameters, even when the Hamiltonian is simple [33]. Pulse finding algorithms also tend to naturally incorporate the physical constraints of the control device, and as such are preferable over naively implemented hard pulses.

Chapter 7

Conclusion

In this thesis we presented the idea of honest error approximations for use in efficient classical simulations of quantum circuits for the purposes of evaluating circuit performance. Strong numerical evidence was provided showing the utility of the method for generating honest representations of circuits under composition. Indeed, the success of the method for errors in the unitary regime is especially surprising and encouraging, given that at the outset of the study, this regime was precisely where underestimation was expected.

As described in Section 5.6, this work has been motivated, from its outset, by the desire for simulations to "honestly" represent the performance of a realistic implementation of a quantum circuit. That is, we want the strength of the errors in the simulation to somehow lower-bound the strength of the errors of the implementation of the circuit we're interested in. Of course, other metrics of simulation performance may be considered as well. For example, in Geller and Zhou's study [36], they start from the viewpoint that it's more desirable for the simulation to be *close* to the performance of the original circuit, and they are happy with some level of underestimation of error strength, so long as the underestimation isn't too bad. In their study, they show that the Pauli twirled error approximations are able to closely model the failure probability of a well-motivated circuit for a variety of error strengths. However, as discussed in Chapter 6, the dominant part of their error model falls into a regime (the "deforming regime") where Pauli twirling the errors doesn't change their character significantly, and therefore the close representation of circuit performance may not be surprising. Given these facts, it would be very interesting to extend their study to include dominant errors in the "unitary regime", where, given the discussion in Chapter 6, orders-of-magnitude underestimation of error strength is possible.

A natural question to ask is whether or not the concept of honest error approximations

could be useful for extending the applicability of threshold theorems. For example, could error models that have so-far eluded threshold theorems be honestly approximated by models that are easier to handle? A related question is how strongly threshold type values depend on the error model. While two error models may be comparable in terms of some measure of error strength (e.g., average gate fidelity or diamond norm distance) they may differ significantly in others. Thus, it is important to understand what measures of individual gate performance best predict the performance of the circuits that they will ultimately be used to implement.

APPENDICES

Appendix A

Mathematical Background

A.1 Spectral Theorem and Singular Value Decomposition

For a finite dimensional Hilbert space $\mathcal{X}$, an operator $A \in \mathrm{L}(\mathcal{X})$ is called *normal* if $A^\dagger A = AA^\dagger$. The most important subset of normal operators for this thesis is $\mathrm{Herm}(\mathcal{X})$.

Theorem A.1.1. *(Spectral Theorem)*
For every normal operator $A \in L(\mathcal{X})$, there exists an orthonormal basis for $\mathcal{X}$, $\{x_1, x_2, ..., x_n\}$, where $n = dim(\mathcal{X})$, of eigenvectors of A, for which

$$A = \sum_{i=1}^{n} \lambda_i x_i x_i^\dagger,$$

where the λ_i are the (not necessarily distinct) eigenvalues of A.

Theorem A.1.2. *(Singular Value Decomposition) For an operator $A \in L(\mathcal{X}, \mathcal{Y})$, there exist orthonormal sets $\{x_1, ..., x_r\} \subset \mathcal{X}$ and $\{y_1, ..., y_r\} \subset \mathcal{Y}$, and positive real numbers $\sigma_1, ..., \sigma_r$ (called the singular values) for which $A = \sum_{i=1}^{r} \sigma_i y_i x_i^\dagger$. The x_i and y_i are eigenvectors of $A^\dagger A$ and $AA^\dagger$ respectively. The singular values are unique, up to ordering, and are the (non-distinct) square roots of the (positive) eigenvalues of $A^\dagger A$ and $AA^\dagger$.*

A.2 Explicit Construction of a Bloch Basis in Arbitrary Finite Dimension

Here we give an explicit construction of a particular "Bloch Basis", following [41], that works in any finite dimension. Let $\mathcal{X} = \mathbb{C}^n$, and let $\{e_k\}_{k=1}^n$ be the standard basis for $\mathcal{X}$. Let $P_{jk} = e_j e_k^\dagger \in \mathrm{L}(\mathcal{X})$, and define the following Hermitian operators

$$W_l = -\frac{1}{\sqrt{l(l+1)}}\left(\sum_{m=1}^{l} P_{mm} - lP_{l+1,l+1}\right) \tag{A.1}$$

$$U_{jk} = \frac{1}{\sqrt{2}}(P_{jk} + P_{kj}) \tag{A.2}$$

$$V_{jk} = \frac{i}{\sqrt{2}}(P_{jk} - P_{kj}), \tag{A.3}$$

for $1 \leq l \leq n-1$ and $1 \leq j < k \leq n$. It can be easily verified that this defines $n^2 - 1$ operators. (Note that these definitions differ from [41] by a factor of $\sqrt{2}$.)

Now, order them as $(W_1, ..., W_{n-1}, U_{12}, ..., U_{n-1,n}, V_{12}, ..., V_{n-1,n})$, and relabel this ordering as $(\lambda_1, ..., \lambda_{n^2-1})$. Finally, define $\lambda_0 = \frac{\mathbb{1}_{\mathcal{X}}}{\sqrt{n}}$. It can be verified that the set of operators $B_{\mathcal{X}} = \{\lambda_i\}_{i=0}^{n^2-1}$ is orthonormal (with respect to the Hilbert-Schmidt inner product), and, given that there are n^2 elements, it is a basis for $\mathrm{L}(\mathcal{X})$.

Appendix B

Partial Ordering Proofs

Here the proofs for the partial ordering results of Chapter 4 are given. The main idea is to convert the partial ordering statements which are statements about all (pure) states into state independent statements that depend only on the matrix representation of channels.

B.1 Quadratic Forms and Schatten 2-norm inequalities

All of what we do in this section is based on the following propositions. The first is a statement about quadratic forms defined on vectorized unit trace Hermitian matrices in the Bloch basis.

Proposition B.1.1. *Let $Q \in \mathbb{R}^{n^2 \times n^2}$ be a symmetric matrix ($Q = Q^T$), and write it in block form as*

$$Q = \left(\begin{array}{c|c} \beta & v^T \\ \hline v & B \end{array}\right),$$

where $\beta \in \mathbb{R}$, $v \in \mathbb{R}^{n^2-1}$, and $B \in \mathbb{R}^{(n^2-1)\times(n^2-1)}$ real matrix. Define $Q_\pm \in \mathbb{R}^{(n^2-1)\times(n^2-1)}$ as

$$Q_\pm = \frac{1}{n}\left[\alpha^2 B + (\beta \pm 2\alpha\|v\|_2)\mathbb{1}_{n^2-1}\right],$$

where $\mathbb{1}_{n^2-1} \in \mathbb{R}^{(n^2-1)\times(n^2-1)}$ *is the identity matrix, and* $\alpha = \sqrt{n-1}$.

Then for any operator $X \in Herm_1(\mathcal{X})$*, satisfying* $\|r\|_2 = 1$ *for* $r = bvec(X)$*, it holds that*

$$r^T Q_- r \leq vec(X)^T Q vec(X) \leq r^T Q_+ r.$$

If $\beta = \|v\|_2 = 0$*, then* $Q_+ = Q_-$*, the above inequalities become equalities, and the statement holds for all* $X \in Herm_1(\mathcal{X})$ *(without restriction on Bloch vector length).*

Proof. We first prove the special case statement when $\beta = \|v\|_2 = 0$. In this case, $Q_\pm = \frac{\alpha^2}{n} B$. Recall that for any $X \in \mathrm{Herm}_1(\mathcal{X})$ with corresponding Bloch vector r, the vectorization of X in the Bloch basis is given by

$$vec(A) = \sqrt{\frac{1}{n}} \left(\begin{array}{c} 1 \\ \hline \alpha r \end{array}\right).$$

It is then clear by direct matrix multiplication that $\mathrm{vec}(X)^T Q \mathrm{vec}(X) = r^T Q_\pm r$ without any assumptions on r.

We now show the general case. Let $X \in \mathrm{Herm}_1(\mathcal{X})$ with Bloch vector r, and we assume $\|r\|_2 = 1$. Then

$$\begin{aligned} \mathrm{vec}(X)^T Q \mathrm{vec}(X) &= \frac{1}{n} \left(\begin{array}{c|c} 1 & \alpha r^T \end{array}\right) \left(\begin{array}{c|c} \beta & v^T \\ \hline v & B \end{array}\right) \left(\begin{array}{c} 1 \\ \hline \alpha r \end{array}\right) \\ &= \frac{1}{n}(\beta + 2\alpha v^T r + \alpha^2 r^T B r). \end{aligned}$$

By the Cauchy-Schwarz inequality, $|v^T r| \leq \|v\|_2 \|r\|_2$, from which follows the following string of inequalities:

$$-\|v\|_2\|r\|_2 \leq -|v^T r| \leq v^T r \leq |v^T r| \leq \|v\|_2 \|r\|_2,$$

and thus

$$\frac{1}{n}(\beta - 2\alpha\|v\|_2\|r\|_2 + \alpha^2 r^T B r) \leq \mathrm{vec}(A)^T Q \mathrm{vec}(A) \leq \frac{1}{n}(\beta + 2\alpha\|v\|_2\|r\|_2 + \alpha^2 r^T B r).$$

By assumption, $\|r\|_2 = 1$, and so

$$\begin{aligned}\frac{1}{n}(\beta \pm 2\alpha\|v\|_2\|r\|_2 + \alpha^2 r^T B r) &= \frac{1}{n}(\beta\|r\|_2^2 \pm 2\alpha\|v\|_2\|r\|_2^2 + \alpha^2 r^T B r) \\ &= r^T Q_\pm r,\end{aligned}$$

from which the desired inequalities follow. □

Corollary B.1.1. *Let $\Lambda_1, \Lambda_2 \in C(\mathcal{X} = \mathbb{C}^n, \mathcal{Y} = \mathbb{C}^m)$, $\Gamma_1, \Gamma_2 \in C(\mathcal{X}, \mathcal{Z} = \mathbb{C}^k)$, and let their Bloch representations be denoted $(M_{\Lambda_1}, t_{\Lambda_1})$, $(M_{\Lambda_2}, t_{\Lambda_2})$, $(M_{\Gamma_1}, t_{\Gamma_1})$ and $(M_{\Gamma_2}, t_{\Gamma_2})$ respectively. Define*

$$\begin{aligned}A = &\frac{m-1}{m}(M_{\Lambda_1} - M_{\Lambda_2})^T(M_{\Lambda_1} - M_{\Lambda_2}) - \frac{k-1}{k}(M_{\Gamma_1} - M_{\Gamma_2})^T(M_{\Gamma_1} - M_{\Gamma_2}) \\ &+ \Bigg(\frac{m-1}{m}\|t_{\Lambda_1} - t_{\Lambda_2}\|_2^2 - \frac{k-1}{k}\|t_{\Gamma_1} - t_{\Gamma_2}\|_2^2 \\ &- 2\left\|\frac{m-1}{m}(M_{\Lambda_1} - M_{\Lambda_2})^T(t_{\Lambda_1} - t_{\Lambda_2}) - \frac{k-1}{k}(M_{\Gamma_1} - M_{\Gamma_2})^T(t_{\Gamma_1} - t_{\Gamma_2})\right\|_2\Bigg)\mathbb{1}_{n^2-1}.\end{aligned}$$

It holds that if $A \geq 0$, then

$$\|\Lambda_1(X) - \Lambda_2(X)\|_2 \geq \|\Gamma_1(X) - \Gamma_2(X)\|_2$$

for all $X \in Herm_1(\mathcal{X})$ satisfying $\|bvec(X)\|_2 = 1$. If both

$$\begin{aligned}\frac{m-1}{m}\|t_{\Lambda_1} - t_{\Lambda_2}\|_2^2 - \frac{k-1}{k}\|t_{\Gamma_1} - t_{\Gamma_2}\|_2^2 &= 0 \\ \left\|\frac{m-1}{m}(M_{\Lambda_1} - M_{\Lambda_2})^T(t_{\Lambda_1} - t_{\Lambda_2}) - \frac{k-1}{k}(M_{\Gamma_1} - M_{\Gamma_2})^T(t_{\Gamma_1} - t_{\Gamma_2})\right\|_2 &= 0,\end{aligned}$$

then the result holds for all $X \in Herm_1(\mathcal{X})$ without restrictions on the length of the Bloch vector.

Proof. Let $S_\mathcal{E}$ denote the matrix representation of the channel $\mathcal{E}$. Let $X \in \mathrm{Herm}_1(\mathcal{X})$. By applying the definition of the 2-norm in terms of the inner product, it is clear that $\|\Lambda_1(X) - \Lambda_2(X)\|_2^2 \geq \|\Gamma_1(X) - \Gamma_2(X)\|_2^2$ is equivalent to $\mathrm{vec}(X)^T Q \mathrm{vec}(X) \geq 0$, where

$$Q = (S_{\Lambda_1} - S_{\Lambda_2})^T(S_{\Lambda_1} - S_{\Lambda_2}) - (S_{\Gamma_1} - S_{\Gamma_2})^T(S_{\Gamma_1} - S_{\Gamma_2}).$$

Now, write the matrix Q in the form of Proposition B.1.1 as

$$Q = \left(\begin{array}{c|c} \beta & v^T \\ \hline v & B \end{array}\right).$$

By direct computation, we can find β, v and B.

$$\begin{aligned}
&(S_{\Lambda_1} - S_{\Lambda_2})^T (S_{\Lambda_1} - S_{\Lambda_2}) \\
&= \left(\begin{array}{c|c} 0 & \sqrt{\frac{n(m-1)}{m}}(t_{\Lambda_1} - t_{\Lambda_2})^T \\ \hline 0 & \sqrt{\frac{n(m-1)}{m(n-1)}}(M_{\Lambda_1} - M_{\Lambda_2})^T \end{array}\right) \left(\begin{array}{c|c} 0 & 0^T \\ \hline \sqrt{\frac{n(m-1)}{m}}(t_{\Lambda_1} - t_{\Lambda_2}) & \sqrt{\frac{n(m-1)}{m(n-1)}}(M_{\Lambda_1} - M_{\Lambda_2}) \end{array}\right) \\
&= \left(\begin{array}{c|c} \frac{n(m-1)}{m}\|t_{\Lambda_1} - t_{\Lambda_2}\|_2^2 & \frac{n(m-1)}{m\sqrt{n-1}}(t_{\Lambda_1} - t_{\Lambda_2})^T (M_{\Lambda_1} - M_{\Lambda_2}) \\ \hline \frac{n(m-1)}{m\sqrt{n-1}}(M_{\Lambda_1} - M_{\Lambda_2})^T (t_{\Lambda_1} - t_{\Lambda_2}) & \frac{n(m-1)}{m(n-1)}(M_{\Lambda_1} - M_{\Lambda_2})^T (M_{\Lambda_1} - M_{\Lambda_2}) \end{array}\right),
\end{aligned}$$

and so, with a similar form for $(S_{\Gamma_1} - S_{\Gamma_2})^T (S_{\Gamma_1} - S_{\Gamma_2})$, we see that the parameters of Q are

$$\begin{aligned}
\beta &= n\left(\frac{m-1}{m}\|t_{\Lambda_1} - t_{\Lambda_2}\|_2^2 - \frac{k-1}{k}\|t_{\Gamma_1} - t_{\Gamma_2}\|_2^2\right) \\
v &= \frac{n}{\sqrt{n-1}}\left(\frac{m-1}{m}(M_{\Lambda_1} - M_{\Lambda_2})^T (t_{\Lambda_1} - t_{\Lambda_2}) - \frac{k-1}{k}(M_{\Gamma_1} - M_{\Gamma_2})^T (t_{\Gamma_1} - t_{\Gamma_2})\right) \\
B &= \frac{n}{n-1}\left(\frac{m-1}{m}(M_{\Lambda_1} - M_{\Lambda_2})^T (M_{\Lambda_1} - M_{\Lambda_2}) - \frac{k-1}{k}(M_{\Gamma_1} - M_{\Gamma_2})^T (M_{\Gamma_1} - M_{\Gamma_2})\right).
\end{aligned}$$

Plugging these expressions for β, v, and B into the expression for Q_- in Proposition B.1.1, reveals that $A = Q_-$ (where A is given in the statement of this corollary). Thus, by Proposition B.1.1, $\mathrm{vec}(X)^T Q \mathrm{vec}(X) \geq \mathrm{bvec}(X)^T A \mathrm{bvec}(X)$ if $\|\mathrm{bvec}(X)\|_2 = 1$, and thus if $A \geq 0$, it follows that

$$\|\Lambda_1(X) - \Lambda_2(X)\|_2^2 - \|\Gamma_1(X) - \Gamma_2(X)\|_2^2 = \mathrm{vec}(X)^T Q \mathrm{vec}(X) \geq 0$$

for $X \in \mathrm{Herm}_1(X)$ satisfying $\|\mathrm{bvec}(X)\|_2 = 1$.

If $\beta = \|v\|_2 = 0$, then it holds for all $X \in \mathrm{Herm}_1(\mathcal{X})$, without restriction on the Bloch vector length, that $\|\Lambda_1(X) - \Lambda_2(X)\|_2^2 - \|\Gamma_1(X) - \Gamma_2(X)\|_2^2 = \mathrm{bvec}(X)^T A \mathrm{bvec}(X)$, and thus $A \geq 0$ implies $\|\Lambda_1(X) - \Lambda_2(X)\|_2 \geq \|\Gamma_1(X) - \Gamma_2(X)\|_2$ for all $X \in \mathrm{Herm}_1(\mathcal{X})$. □

B.2 Relating Norm Inequalities

Lemma B.2.1. *Let $A \in Herm(\mathcal{X})$ such that $Tr(A) = 0$ and $rank(A) \leq 2$. Then it holds that $\|A\|_p = 2^{\frac{1}{p}-1}\|A\|_1$.*

Proof. $\mathrm{Tr}(A) = 0$ and $\mathrm{rank}(A) \leq 2$ implies that A has at most two non-zero eigenvalues of equal magnitude and opposite sign $\{\pm\lambda\}$. Then, by definition, $\|A\|_p = (2|\lambda|^p)^{\frac{1}{p}} = 2^{\frac{1}{p}}|\lambda| = 2^{\frac{1}{p}-1}\|A\|_1$. □

Lemma B.2.2. *Let $A \in Herm(\mathcal{X})$ and $B \in Herm(\mathcal{Y})$ such that $Tr(A) = Tr(B) = 0$ and $rank(B) \leq 2$, and let $p \geq 1$. Then if $\|A\|_p \geq \|B\|_p$, it holds that $\|A\|_1 \geq \|B\|_1$.*

Proof. First we show that for any traceless Hermitian operator A, $\|A\|_1 \geq 2^{\frac{1}{p}-1}\|A\|_p$. Let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be two positive semi-definite operators such that $A = P - Q$ and $PQ = 0$. Then, $\mathrm{Tr}(A) = 0$ means that $\|P\|_1 = \mathrm{Tr}(P) = \mathrm{Tr}(Q) = \|Q\|_1$, and as $PQ = 0$, we have that $\|A\|_p^p = \|P\|_p^p + \|Q\|_p^p$. Thus

$$\|A\|_p^p = \|P\|_p^p + \|Q\|_p^p \leq \|P\|_1^p + \|Q\|_1^p = 2\|P\|_1^p = 2(\frac{1}{2}\|A\|_1)^p = 2^{1-p}\|A\|_1^p,$$

where the inequality follows from $\|X\|_p \leq \|X\|_1$ for *any* operator X and $p \geq 1$.

From Lemma B.2.1, by the assumptions on B, it follows that $\|B\|_1 = 2^{\frac{1}{p}-1}\|B\|_p$, and thus it follows that $\|A\|_1 \geq 2^{\frac{1}{p}-1}\|A\|_p \geq 2^{\frac{1}{p}-1}\|B\|_p = \|B\|_1$. □

B.3 Partial Ordering of Quantum Channels

We can now apply the statements of the previous sections to situations of interest regarding quantum channels. These are the proofs to the Corollaries in Section 4.

Proof of Corollary 4.1.1. This is a straightforward application of Corollary B.1.1. In the notation of that Corollary, make the associations $\Lambda_1 = \Gamma_1 = \mathbb{1}_{\mathrm{L}(\mathcal{X})}$ (which has Bloch representation $(\mathbb{1}_{n^2-1}, 0)$), and $\Lambda_2 = \Lambda$, $\Gamma_2 = \Gamma$, with corresponding Bloch representations (M_Λ, t_Λ) and (M_Γ, t_Γ). The conditions of this Corollary can be obtained by plugging these Bloch representations into the conditions in Corollary B.1.1 (and removing dimensional factors), to obtain the desired result. □

Proof of Corollary 4.1.2. If $\Lambda \geq_p^* \Gamma$, then for any pure $\rho \in \mathrm{D}(\mathcal{X})$, $\|\rho - \Lambda(\rho)\|_p \geq \|\rho - \Gamma(\rho)\|_p$. By the assumption that $\mathrm{rank}(\rho - \Gamma(\rho)) \leq 2$ for all pure states ρ, and by the fact that $\rho - \Gamma(\rho) \in \mathrm{Herm}(\mathcal{X})$ and $\mathrm{Tr}(\rho - \Gamma(\rho)) = 0$, Lemma B.2.2 then gives us that $\|\rho - \Lambda(\rho)\|_1 \geq \|\rho - \Gamma(\rho)\|_1$ for all pure ρ, ie. $\Lambda \geq_1^* \Gamma$. □

Proof of Corollary 4.1.3. By Lemma B.2.1, for any $\rho_1, \rho_2 \in \mathrm{D}(\mathcal{X})$, $\|\rho_1 - \rho_2\|_q = 2^{\frac{1}{q}-\frac{1}{p}}\|\rho_1 - \rho_2\|_p$, thus any inequality on differences of qubit states with the Schatten q-norm can be converted into one of the Schatten p-norm. □

Before proving Corollary 4.1.4, we first prove the following proposition.

Proposition B.3.1. *Let $\Lambda, \Gamma \in C(\mathcal{X})$, $p \geq 1$, $A \in L(\mathcal{X}, \mathcal{X} \otimes \mathcal{Z})$ be a Stinespring representation of Γ for some space $\mathcal{Z}$, and $u \in \mathcal{Z}$ an arbitrary unit vector. Define the maps $\Gamma_1, \Gamma_2 \in C(\mathcal{X}, \mathcal{X} \otimes \mathcal{Z})$ as $\Gamma_1(X) = X \otimes uu^\dagger$ and $\Gamma_2(X) = AXA^\dagger$ for $X \in L(\mathcal{X})$. If*

$$\|\rho - \Lambda(\rho)\|_p \geq \|\Gamma_1(\rho) - \Gamma_2(\rho)\|_p$$

for all pure states $\rho \in D(\mathcal{X})$, it follows that $\Lambda \geq_1^* \Gamma$ and $\Gamma \geq_F^* \Lambda$.

Proof. The important thing to note is that both Γ_1 and Γ_2 map pure states in $\mathrm{D}(\mathcal{X})$ to pure states in $\mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$. Thus, for any pure state $\rho \in \mathrm{D}(\mathcal{X})$, it follows that $\mathrm{rank}(\Gamma_1(\rho) - \Gamma_2(\rho)) \leq 2$, and we can then apply Lemma B.2.2 to conclude that $\|\rho - \Lambda(\rho)\|_1 \geq \|\Gamma_1(\rho) - \Gamma_2(\rho)\|_1$ for all pure states $\rho \in \mathrm{D}(\mathcal{X})$. From the fact that $\|\cdot\|_1$ is non-increasing under partial trace ($\|A\|_1 \geq \|\mathrm{Tr}_{\mathcal{Z}}(A)\|_1$ for all $A \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Z})$), it follows that $\Lambda \geq_1^* \Gamma$.

To get that $\Gamma \geq_F^* \Lambda$, we need one of the Fuchs-van de Graaf inequalities, $F(\rho_1, \rho_2) \leq \sqrt{1 - \frac{1}{4}\|\rho_1 - \rho_2\|_1^2}$, that there is equality if ρ_1, ρ_2 are pure, and Uhlmann's Theorem, which in particular implies that $F(\rho_1, \rho_2) \geq F(u_1 u_1^\dagger, u_2 u_2^\dagger)$ for any purifications u_1 and u_2 of ρ_1 and ρ_2 respectively. The first two facts (along with the preceding part of the proof) give that, for any pure state $\rho \in \mathrm{D}(\mathcal{X})$,

$$F(\Lambda(\rho), \rho) \leq \sqrt{1 - \frac{1}{4}\|\rho - \Lambda(\rho)\|_1^2} \leq \sqrt{1 - \frac{1}{4}\|\Gamma_1(\rho) - \Gamma_2(\rho)\|_1^2} = F(\Gamma_1(\rho), \Gamma_2(\rho)).$$

As $\Gamma_1(\rho)$ and $\Gamma_2(\rho)$ are purifications of ρ and $\Gamma(\rho)$ respectively, it follows that $F(\Gamma_1(\rho), \Gamma_2(\rho)) \leq F(\rho, \Gamma(\rho))$, and therefore, as ρ was an arbitrary pure state, it follows that $\Gamma \geq_F^* \Lambda$. □

Proof of Corollary 4.1.4. By Corollary B.1.1, if $B \geq 0$, then $\|\rho - \Lambda(\rho)\|_2 \geq \|\Gamma_1(\rho) - \Gamma_2(\rho)\|_2$ for all pure $\rho \in \mathrm{D}(\mathcal{X})$. By Proposition B.3.1, the result follows. □

Appendix C

Sampled Estimation of Parameters

This appendix gives a brief analysis of how accurately the hedging parameters for honest approximations, defined in Section 4.3, are estimated by sampling $N = 10^6$ points (a number which is used throughout this thesis). Recall, for two channels $\Lambda, \Gamma \in \mathrm{C}(\mathcal{X})$, and a state $\rho \in \mathrm{D}(\mathcal{X})$ we defined the *hedging* of Λ with respect to Γ for the state ρ as:

$$h(\Lambda, \Gamma, \rho) = \|\rho - \Lambda(\rho)\|_1 - \|\rho - \Gamma(\rho)\|_1. \tag{C.1}$$

We are interested in two quantities: the average of h over all pure states, which we denote $\bar{h}(\Lambda, \Gamma)$, and the probability that any given pure state produces negative hedging, which we denote p_{viol}.

More formally, $\bar{h}(\Lambda, \Gamma)$ is the mean of the random variable $H \equiv h(\Lambda, \Gamma, |\psi\rangle \langle\psi|)$, where $|\psi\rangle$ a random variable distributed according to the uniform measure on states (induced by the Haar measure), and p_{viol} is the probability, $p = \mathrm{Prob}(P = 1)$, where P is defined as the random variable $P \equiv f\left(h(\Lambda, \Gamma, |\psi\rangle \langle\psi|)\right)$, where $f(x) = 1$ if $x < 0$ and $f(x) = 0$ if $x \geq 0$ (ie. the function which indicates if its argument is negative), and $|\psi\rangle$ is again distributed uniformly (note that p is also the expectation value of P). By naive application of Chebyshev's inequality, we can see that uniformly sampling $N = 10^6$ pure states provides good estimates for these quantities.

Recall, Chebyshev's inequality gives that, for a random variable X with mean μ and variance σ^2, the probability that the average of N independent realizations of X, $\bar{X}$, deviates from the mean by more than $a > 0$ is bounded by

$$\mathrm{Prob}\left(|\bar{X} - \mu| \geq a\right) \leq \frac{\sigma^2}{Na^2}. \tag{C.2}$$

By applying the definition of H, we can compute a very loose upper bound on its variance, σ_H^2 (with its mean denoted μ_H):

$$\begin{aligned}
\sigma_H^2 &= \mathbb{E}[H^2] - \mu_H^2 && \text{(C.3)}\\
&\leq \mathbb{E}[H^2] && \text{(C.4)}\\
&= \int d\psi h(\Lambda, \Gamma, |\psi\rangle\langle\psi|)^2 && \text{(C.5)}\\
&= \int d\psi \left(\| |\psi\rangle\langle\psi| - \Lambda(|\psi\rangle\langle\psi|) \|_1 - \| |\psi\rangle\langle\psi| - \Gamma(|\psi\rangle\langle\psi|) \|_1 \right)^2 && \text{(C.6)}\\
&\leq \int d\psi \, \| \Lambda(|\psi\rangle\langle\psi|) - \Gamma(|\psi\rangle\langle\psi|) \|_1^2 && \text{(C.7)}\\
&\leq \int d\psi \, \| \Lambda - \Gamma \|_\diamond^2 && \text{(C.8)}\\
&= \| \Lambda - \Gamma \|_\diamond^2 , && \text{(C.9)}
\end{aligned}$$

where the second last inequality follows from the triangle inequality, the last inequality from the definition of $\| \cdot \|_\diamond$ (see Section 3), and the last equality by $\int d\psi = 1$. Note that this is likely a very poor bound on σ_H^2, but it will be sufficient for our purposes.

Thus, for $\bar{H} = \frac{1}{N}(H_1 + ... + H_N)$, where H_i is independently drawn according to H, for any $a > 0$,

$$\text{Prob}\left(|\bar{H} - \mu_H| \geq a \right) \leq \frac{\| \Lambda - \Gamma \|_\diamond^2}{N a^2}. \tag{C.10}$$

Letting $a = 10^{-2} \| \Lambda - \Gamma \|_\diamond$ and $N = 10^6$, we arrive at the expression:

$$\text{Prob}\left(|\bar{H} - \mu_H| \geq 10^{-2} \| \Lambda - \Gamma \|_\diamond \right) \leq 10^{-2}. \tag{C.11}$$

In other words, there is at most a 1% chance that our estimate of μ_H differs from the true value by more than $10^{-2} \| \Lambda - \Gamma \|_\diamond$. In the vast majority of cases within this thesis, the computed value of $\bar{h}$ is on the same order of magnitude as $\| \Lambda - \Gamma \|_\diamond$, and as such, in these cases, it is unlikely that our estimate of μ_H deviates from the true value in any significant way. In a few cases, the estimate of μ_H is on the same order of magnitude as $10^{-2} \| \Lambda - \Gamma \|_\diamond$, and some are as low as $10^{-4} \| \Lambda - \Gamma \|_\diamond$. In the former case, our bound gives a good chance that we have the first significant digit right (ie. a good order-of-magnitude estimate) and in the latter case, our bound isn't of much use. Nevertheless, given that this bound was derived in an incredibly loose way, and given that, in these cases, the hedging

is so small as to be considered negligible, we can conclude that the estimates still provide good qualitative insight.[1]

To examine our computation of p_{viol}, note again that we're trying to estimate $p = \text{Prob}(P = 1) = \mathbb{E}[P]$. In such a case, $\sigma_P^2 = \mathbb{E}\left[(P-p)^2\right] = (1-p)^2 p + (0-p)^2(1-p) \leq 1/4$. Thus, applying Chebyshev's inequality gives that, for $\bar{P} = \frac{1}{N}(P_1 + ... + P_N)$, where P_i are independent realizations of P,

$$\text{Prob}(|\bar{P} - \mu_P| \geq 10^{-2}) \leq (1/4) \times 10^{-2}, \tag{C.12}$$

when $N = 10^6$ samples are used. In other words, our computed values of p_{viol} have less than a 0.25% chance of being off from the true value by more than 1%, which is more than acceptable for our purposes.

[1] The most important estimates of this parameter are given in Table 5.2 in Section 5.5, where we test how well our approximations represent circuit performance. The average hedging is the same order of magnitude as the diamond norm in all but two cases – both for Pauli twirled approximations. In one case it is one order of magnitude smaller, and thus our computed value is still very good. In the other case, it is three orders of magnitude smaller, and thus there is, roughly, at most a 10% that our computed quantity differs in the most significant digit. Again, given that our bounds on this probability are derived naively and very loosely, we are confident that this is an accurate estimate.

Appendix D

Gate Statistics

Table D.1: Statistics for the various approximations of the gates in GS1, approximated from $N = 10^6$ random pure states.

GS1					
Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
$\mathbb{1}$	χ_{00}	0.999994	0.999994	0.997618	0.998314
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.76×10^{-3}	1.20×10^{-5}	4.76×10^{-3}	4.77×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.76×10^{-3}	6.73×10^{-3}	3.64×10^{-3}
	$\overline{h}$		-3.73×10^{-3}	1.14×10^{-7}	1.64×10^{-6}
	p_{viol}		1.	0.	0.
X	χ_{00}	0.996147	0.996147	0.991234	0.994179
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.51×10^{-2}	7.71×10^{-3}	1.75×10^{-2}	1.59×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.29×10^{-2}	1.67×10^{-2}	4.58×10^{-3}
	$\overline{h}$		-5.76×10^{-3}	1.12×10^{-3}	3.90×10^{-4}
	p_{viol}		0.99892	0.	0.
Y	χ_{00}	0.998330	0.998330	0.992021	0.994711
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.35×10^{-2}	3.34×10^{-3}	1.60×10^{-2}	1.43×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.30×10^{-2}	1.86×10^{-2}	8.15×10^{-3}
	$\overline{h}$		-7.96×10^{-3}	9.94×10^{-4}	3.09×10^{-4}
	p_{viol}		0.99953	0.	0.
Z	χ_{00}	0.995289	0.995289	0.992055	0.993663
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.06×10^{-2}	9.42×10^{-3}	1.59×10^{-2}	1.31×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.59×10^{-3}	7.68×10^{-3}	3.46×10^{-3}
	$\overline{h}$		-5.71×10^{-4}	3.01×10^{-3}	1.29×10^{-3}
	p_{viol}		0.60369	0.	0.
H	χ_{00}	0.997642	0.997642	0.991031	0.993638
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.35×10^{-2}	4.72×10^{-3}	1.79×10^{-2}	1.54×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.27×10^{-2}	1.85×10^{-2}	8.83×10^{-3}
	$\overline{h}$		-6.42×10^{-3}	2.51×10^{-3}	1.10×10^{-3}
	p_{viol}		1.	0.	0.
CNOT	χ_{00}	0.993152	0.993152	0.976600	0.976600
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.09×10^{-2}	1.37×10^{-2}	4.68×10^{-2}	4.68×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.82×10^{-2}	4.52×10^{-2}	4.52×10^{-2}
	$\overline{h}$		-1.17×10^{-2}	1.52×10^{-2}	1.52×10^{-2}
	p_{viol}		1.	0.	0.

Table D.2: Statistics for the various approximations of the gates in GS2, approximated from $N = 10^6$ random pure states.

GS2					
Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
$\mathbb{1}$	χ_{00}	0.999087	0.999087	0.999085	0.999086
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.48×10^{-5}	2.50×10^{-5}	2.06×10^{-6}
	$\overline{h}$		-1.63×10^{-7}	2.65×10^{-6}	8.34×10^{-7}
	p_{viol}		0.49861	0.	0.
X	χ_{00}	0.999074	0.999074	0.998972	0.999029
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.91×10^{-3}	1.85×10^{-3}	2.06×10^{-3}	2.00×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.38×10^{-4}	4.85×10^{-4}	1.16×10^{-4}
	$\overline{h}$		-4.04×10^{-5}	9.61×10^{-5}	6.12×10^{-5}
	p_{viol}		0.7897	0.	0.
Y	χ_{00}	0.999075	0.999075	0.998990	0.999033
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.91×10^{-3}	1.85×10^{-3}	2.02×10^{-3}	1.99×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.32×10^{-4}	4.72×10^{-4}	1.03×10^{-4}
	$\overline{h}$		-4.02×10^{-5}	7.32×10^{-5}	5.58×10^{-5}
	p_{viol}		0.8674	0.	0.
Z	χ_{00}	0.999023	0.999023	0.998869	0.998901
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.97×10^{-3}	1.95×10^{-3}	2.26×10^{-3}	2.21×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.67×10^{-4}	4.27×10^{-4}	3.33×10^{-4}
	$\overline{h}$		-1.37×10^{-5}	1.87×10^{-4}	1.52×10^{-4}
	p_{viol}		0.61392	0.	0.
H	χ_{00}	0.999060	0.999060	0.998684	0.998878
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	2.33×10^{-3}	1.88×10^{-3}	2.63×10^{-3}	2.53×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.25×10^{-3}	1.45×10^{-3}	4.35×10^{-4}
	$\overline{h}$		-3.05×10^{-4}	1.98×10^{-4}	1.32×10^{-4}
	p_{viol}		1.	0.	0.
CNOT	χ_{00}	0.998071	0.998071	0.997683	0.997683
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.99×10^{-3}	3.86×10^{-3}	4.63×10^{-3}	4.63×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.01×10^{-3}	1.33×10^{-3}	1.33×10^{-3}
	$\overline{h}$		-7.38×10^{-5}	5.44×10^{-4}	5.44×10^{-4}
	p_{viol}		0.76905	0.	0.

Table D.3: Statistics for the various approximations of the gates in GS3, approximated from $N = 10^6$ random pure states.

GS3					
Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
$\mathbb{1}$	χ_{00}	0.998751	0.998751	0.996501	0.996501
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.99×10^{-3}	2.50×10^{-3}	7.00×10^{-3}	7.00×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.50×10^{-3}	5.28×10^{-3}	5.28×10^{-3}
	$\overline{h}$		-1.03×10^{-3}	1.91×10^{-3}	1.91×10^{-3}
	p_{viol}		0.74978	0.	0.
X	χ_{00}	0.998751	0.998751	0.997447	0.997674
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.17×10^{-3}	2.50×10^{-3}	5.11×10^{-3}	4.65×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.37×10^{-3}	3.12×10^{-3}	2.75×10^{-3}
	$\overline{h}$		-3.64×10^{-4}	1.39×10^{-3}	1.06×10^{-3}
	p_{viol}		0.69216	0.	0.
Y	χ_{00}	0.998751	0.998751	0.997669	0.997669
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.17×10^{-3}	2.50×10^{-3}	4.66×10^{-3}	4.66×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.38×10^{-3}	2.76×10^{-3}	2.76×10^{-3}
	$\overline{h}$		-3.63×10^{-4}	1.06×10^{-3}	1.06×10^{-3}
	p_{viol}		0.69113	0.	0.
Z	χ_{00}	0.998751	0.998751	0.996501	0.996501
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.99×10^{-3}	2.50×10^{-3}	7.00×10^{-3}	7.00×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.50×10^{-3}	5.28×10^{-3}	5.28×10^{-3}
	$\overline{h}$		-1.03×10^{-3}	1.92×10^{-3}	1.91×10^{-3}
	p_{viol}		0.75055	0.	0.
H	χ_{00}	0.998751	0.998751	0.996883	0.996940
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.31×10^{-3}	2.50×10^{-3}	6.23×10^{-3}	6.12×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.27×10^{-3}	4.43×10^{-3}	4.36×10^{-3}
	$\overline{h}$		-7.92×10^{-4}	1.68×10^{-3}	1.60×10^{-3}
	p_{viol}		0.74018	0.	0.
CNOT	χ_{00}	0.997441	0.997441	0.981266	0.981266
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.98×10^{-2}	5.12×10^{-3}	3.75×10^{-2}	3.75×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.83×10^{-2}	3.90×10^{-2}	3.90×10^{-2}
	$\overline{h}$		-1.15×10^{-2}	1.48×10^{-2}	1.48×10^{-2}
	p_{viol}		0.99988	0.	0.

Appendix E

Physical Models and Gate Protocol Details

This is a reproduction of Appendix B from [12]. It should be considered technically separate from this thesis, and is only included for completeness.

This appendix describes the physical models, noise refocusing techniques, and the subsequent gate sets generated from these. It is assumed that the density matrix $\rho(t)$ describing a physical system evolves according to

$$\frac{\partial}{\partial t}\rho(t) = -i[H(t), \rho(t)] + \sum_i \left(L_i \rho(t) L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho(t)\} \right), \tag{E.1}$$

where $H(t)$ is the Hamiltonian for the system and $\{L_i\}$ is a set of Lindblad operators generating non-unitary dynamics [42, 43, 44]. A physical model must specify all deterministic and stochastic Hamiltonians (both internal, and control), and specify any Lindblad operators that the system is subject to.

E.1 Physical Model 1

Physical Model 1 is motivated by a double quantum dot physical system. A double quantum dot is a pair of electrons contained in a double potential well. The spatial and spin states of the electrons encode logical states $|0\rangle$ and $|1\rangle$,

$$|0\rangle = |\Phi_{11}^T\rangle \otimes (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/\sqrt{2} \tag{E.2}$$

$$|1\rangle = \left(a\,|\Phi_{11}^S\rangle + b\,|\Phi_{02}^S\rangle\right) \otimes (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}, \tag{E.3}$$

where $|\Phi_{11}^S\rangle$ and $|\Phi_{11}^T\rangle$ are symmetric and anti-symmetric spatial states with one electron in each of the potential wells, and $|\Phi_{02}^S\rangle$ is a symmetric spatial state having two electrons in one particular well.

The electron state can be controlled by varying the voltage detuning $B(t)$ and Zeeman splitting difference $A(t)$, described below.

- **Voltage detuning** introduces a potential energy difference B between the quantum wells. $B > 0$ favours the $|\Phi_{02}^S\rangle$ spatial state over $|\Phi_{11}^S\rangle$ and $|\Phi_{11}^T\rangle$, because $|\Phi_{02}^S\rangle$ allows for both electrons to minimize their potential energy. The parameters a and b in Equation (E.3) are therefore B-dependent and given by Fermi-Dirac statistics [45] such that the probability $p_{11} = |a|^2$ of having an electron with potential energy B is given by $p_{11} = \frac{1}{1+e^{B/B_1-B_2}}$, whereby

$$a(B) = \sqrt{\frac{1}{1+e^{B/B_1-B_2}}}$$
$$b(B) = \sqrt{\frac{1}{1+e^{-(B/B_1-B_2)}}}.$$

  The detuning Hamiltonian $H(B)$ is diagonal in the spatial states $\{|\Phi_{11}^S\rangle, |\Phi_{11}^T\rangle, |\Phi_{02}^S\rangle\}$ and takes a form

$$H(B) = B\left(|\Phi_{11}^S\rangle\langle\Phi_{11}^S| + |\Phi_{11}^T\rangle\langle\Phi_{11}^T|\right) + B_0\,|\Phi_{02}^S\rangle\langle\Phi_{02}^S|, \tag{E.4}$$

  where B_0 is the energy eigenvalue of $|\Phi_{02}\rangle$ at zero detuning. Up to a constant identity contribution, this yields the following Hamiltonian for logical states

$$H(B) = \begin{pmatrix} \langle 0|\,H(B)\,|0\rangle & \langle 0|\,H(B)\,|1\rangle \\ \langle 1|\,H(B)\,|0\rangle & \langle 1|\,H(B)\,|1\rangle \end{pmatrix} = \frac{1}{2}\begin{pmatrix} |b|^2(B-B_0) & 0 \\ 0 & -|b|^2(B-B_0) \end{pmatrix}.$$

- **Zeeman splitting** is related to the energy difference between electron spin-up and spin-down states in the presence of an external magnetic field. A magnetic field gradient across the potential wells introduces an energy splitting A between $|\Phi_{11}\rangle \otimes |\uparrow\downarrow\rangle$ and $|\Phi_{11}\rangle \otimes |\downarrow\uparrow\rangle$, where $|\Phi_{11}\rangle = |\Phi_{11}^S\rangle + |\Phi_{11}^T\rangle$, leading to a Hamiltonian $H(A) = \frac{A}{2}\left(|\Phi_{11}\rangle\langle\Phi_{11}| \otimes |\uparrow\downarrow\rangle\langle\uparrow\downarrow| - |\Phi_{11}\rangle\langle\Phi_{11}| \otimes |\downarrow\uparrow\rangle\langle\downarrow\uparrow|\right)$. The resulting matrix $H(A)$ in the logical basis is

$$H(A) = \begin{pmatrix} \langle 0|\,H(A)\,|0\rangle & \langle 0|\,H(A)\,|1\rangle \\ \langle 1|\,H(A)\,|0\rangle & \langle 1|\,H(A)\,|1\rangle \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & aA \\ a^*A & 0 \end{pmatrix}.$$

Substituting the parameters $a(B)$ and $b(B)$ into the above expressions and summing them results in the effective logical single qubit Hamiltonian

$$H(t) = \frac{1}{2}\frac{A(t)+\alpha(t)}{\sqrt{1+\exp\left(\frac{B(t)}{B_1}-B_2\right)}}X + \frac{1}{2}\frac{B(t)-B_0+\beta(t)}{1+\exp\left[-\left(\frac{B(t)}{B_1}-B_2\right)\right]}Z, \tag{E.5}$$

where the parameters $\alpha(t)$ and $\beta(t)$ encapsulate the stochastic behaviour of the parameters $A(t)$ and $B(t)$. As the logical state $|0\rangle$ corresponds to the ground state of the physical Hamiltonian at zero detuning ($B = 0$) and in zero magnetic field gradient ($A = 0$), we make an assumption that relaxation acts on the logical state simply as a Lindblad operator $L = \frac{1}{2\sqrt{T_1}}(X + iY)$.

The parameters $A(t)$, $B(t)$, B_i for $i \in \{0, 1, 2\}$, and T_1, define the deterministic evolution of the system, with the first two representing the single-qubit controls and the latter being the T_1 time-constant of the system. Both control scalars are specified for intervals of length δt, over which the controls remain constant, whereas the rate of change of these control scalars between adjacent intervals is bounded with $\left|\frac{dA(t)}{dt}\right| \leq \Delta A_{\max}$ and $\left|\frac{dB(t)}{dt}\right| \leq \Delta B_{\max}$. Additionally, the controls are bounded by some maximum value; that is $|A(t)| \leq A_{\max}$ and $0 \leq B(t) \leq B_{\max}$ for some $A_{\max}, B_{\max} \geq 0$.

The parameters specified by α and β are independent, zero-mean, stationary Gaussian processes [46], such that $\langle\alpha(t)\rangle = \langle\beta(t)\rangle = 0$ and $\langle\alpha(t_1)\beta(t_2)\rangle = 0$. The auto-correlation functions are given by

$$\begin{aligned}\langle\alpha(t_1)\alpha(t_2)\rangle &= \Gamma_{\alpha_1}^2\ \delta(|t_1-t_2|) + \Gamma_{\alpha_2}^2 e^{-\left(\frac{|t_1-t_2|}{\tau_1}\right)^2+\left(\frac{|t_1-t_2|}{\tau_2}\right)^4-\left(\frac{|t_1-t_2|}{\tau_3}\right)^6} \\ \langle\beta(t_1)\beta(t_2)\rangle &= \Gamma_{\beta_1}^2 + \Gamma_{\beta_2}^2\ \delta(|t_1-t_2|),\end{aligned} \tag{E.6}$$

where $\delta(t)$ is the Dirac delta function. Parameters labelled with the letter Γ represent the noise strengths, and those labelled with τ represent various correlation times.

The Hamiltonian for simulating two-qubit gates is given by

$$\begin{aligned}H(t) &= H^{(1)}(t)\otimes\mathbb{1} + \mathbb{1}\otimes H^{(2)}(t) + H_{zz}(t) \\ H_{zz}(t) &= \frac{1}{4}\frac{C(t)(1+\gamma(t))}{\left(1+\exp\left[-\left(\frac{B^{(1)}(t)}{B_1}-B_2\right)\right]\right)} \times \frac{Z\otimes Z - Z\otimes\mathbb{1} - \mathbb{1}\otimes Z}{\left(1+\exp\left[-\left(\frac{B^{(2)}(t)}{B_1}-B_2\right)\right]\right)},\end{aligned}$$

with two Lindblad operators $L_1 = \frac{1}{2\sqrt{T_1}}(X + iY)\otimes\mathbb{1}$ and $L_2 = \frac{1}{2\sqrt{T_1}}\mathbb{1}\otimes(X + iY)$. Any parameters or Hamiltonians denoted by superscript (i) mark either the first ($i = 1$) or

the second ($i = 2$) qubit, and are identical to the Hamiltonian in line (E.5). The stochastic parts for single-qubit Hamiltonians on different qubits are taken to be independent. The two-qubit control parameter $C(t)$ can only take two values, $C(t) \in \{0, C_{\max}\}$, and the noise parameter $\gamma(t)$ is an independent zero-mean stationary Gaussian process with autocorrelation function

$$\langle \gamma(t_1)\gamma(t_2)\rangle = \Gamma_\gamma^2 \delta(|t_1 - t_2|). \tag{E.7}$$

E.2 Physical Model 2

Physical Model 2 is an archetypal two level system. For a single qubit, the Hamiltonian is given by

$$H(t) = \frac{1}{2}[B(t)(1 + \beta_1(t)) + \beta_2(t)]Z + \frac{1}{2}A(t)(1 + \alpha(t))\left[\cos(\phi(t))X + \sin(\phi(t))Y\right], \tag{E.8}$$

with the only Lindblad operator given by $L = \frac{1}{2\sqrt{T_1}}(X + iY)$.

The parameters $A(t)$, $B(t)$, $\phi(t)$, and T_1, define the deterministic evolution of the system, with the first three representing the single-qubit controls. Every control value is specified for intervals of length δt, over which the controls remain constant. Each control scalar is bounded by some maximum value; that is $|A(t)| \leq A_{\max}$ and $|B(t)| \leq B_{\max}$ for some $A_{\max}, B_{\max} \geq 0$, but is not limited by any control rates.

The parameters specified using the letters α and β are all stationary Gaussian processes. All are zero-mean and independent. That is, $\langle \alpha(t)\rangle = \langle \beta_i(t)\rangle = 0$ for $i = 1, 2$, and $\langle \alpha(t_1)\beta_i(t_2)\rangle = \langle \beta_1(t_1)\beta_2(t_2)\rangle = 0$ for $i = 1, 2$. The auto-correlation functions are given as

$$\langle \alpha(t_1)\alpha(t_2)\rangle = \Gamma_\alpha^2 \; g_{1/f}(\Lambda_\alpha^{(l)}, \Lambda_\alpha^{(u)}, |t_1 - t_2|) \tag{E.9}$$

$$\langle \beta_1(t_1)\beta_1(t_2)\rangle = \Gamma_{\beta_1}^2 \; g_{1/f}(\Lambda_{\beta_1}^{(l)}, \Lambda_{\beta_1}^{(u)}, |t_1 - t_2|) \tag{E.10}$$

$$\langle \beta_2(t_1)\beta_2(t_2)\rangle = \Gamma_{\beta_2}^2 \; g_{1/f}(\Lambda_{\beta_2}^{(l)}, \Lambda_{\beta_2}^{(u)}, |t_1 - t_2|), \tag{E.11}$$

where the parameters labelled with the letter Γ are the noise strengths and those labelled with Λ represent upper and lower cutoffs for $1/f$ noise. The autocorrelation function $g_{1/f}$ for $1/f$ noise is defined as the Fourier transform of $1/f$ spectral density with smooth cutoffs [47]

$$g_{1/f}(\Lambda_1, \Lambda_2, \Delta t) = \int_{-\infty}^{\infty} \frac{2}{\pi\omega}\left(\arctan\left(\frac{\omega}{\Lambda_1}\right) - \arctan\left(\frac{\omega}{\Lambda_2}\right)\right) e^{-i\omega\Delta t}\, \mathrm{d}\omega. \tag{E.12}$$

Notice that $\lim_{\Lambda_1 \to 0, \Lambda_2 \to \infty} \frac{2}{\pi\omega} \left(\arctan\left(\frac{\omega}{\Lambda_1}\right) - \arctan\left(\frac{\omega}{\Lambda_2}\right)\right) = \frac{1}{|\omega|}$.

The two-qubit Hamiltonian for this model is given by

$$\begin{aligned} H(t) &= H^{(1)}(t) \otimes \mathbb{1} + \mathbb{1} \otimes H^{(2)}(t) + H_{zz}(t) \\ H_{zz}(t) &= -\frac{1}{2} C(t)(1 + \gamma(t)) Z \otimes Z, \end{aligned} \tag{E.13}$$

with two Lindblad operators $L_1 = \frac{1}{2\sqrt{T_1}}(X + iY) \otimes \mathbb{1}$ and $L_2 = \frac{1}{2\sqrt{T_1}} \mathbb{1} \otimes (X + iY)$.

Single-qubit Hamiltonians denoted by $H^{(i)}$ acting either on the first ($i = 1$) or the second ($i = 2$) qubit have identical parameters to the Hamiltonian in Equation (E.8), and stochastic parts for single-qubit Hamiltonians are taken to be independent. The two-qubit control parameter $C(t)$ is bounded in its maximum value $|C(t)| \leq C_{\max}$, but is otherwise unconstrained. $\gamma(t)$ is an independent zero-mean stationary Gaussian process, its autocorrelation function being given by

$$\langle \gamma(t_1)\gamma(t_2) \rangle = \Gamma_\gamma^2 \; g_{1/f}(\Lambda_\gamma^{(l)}, \Lambda_\gamma^{(u)}, |t_1 - t_2|). \tag{E.14}$$

E.3 XY Sequence Gate Protocol

Suppose we have a dynamical decoupling sequence which is given as a list of unitary operations $\{A_i\}$, $i = 1, ..., N$, where i denotes the temporal order of these operations. We demand that

$$A_N \; ... \; A_2 A_1 = e^{i\phi} \; \mathbb{1}, \tag{E.15}$$

where $e^{i\phi}$ is an arbitrary global phase. If we want to spread a unitary gate U across the sequence $\{A_i\}$, we first find

$$U_1 = A_1 U^{1/N} A_1^{-1} \tag{E.16}$$

$$U_2 = A_2 A_1 U^{1/N} A_1^{-1} A_2^{-1} \tag{E.17}$$

$$\vdots \tag{E.18}$$

$$U_N = A_N \; ... \; A_2 A_1 U^{1/N} A_1^{-1} A_2^{-1} \; ... \; A_N^{-1}, \tag{E.19}$$

where $\left(U^{1/N}\right)^N = U$, and then implement the sequence A_1, U_1, A_2, U_2,..., A_N, U_N resulting in $U_N A_N \; ... \; U_2 A_2 U_1 A_1 = e^{i\phi} U$, which follows from direct substitution and Equation (E.15).

E.3.1 XY8 sequence

The XY8 sequence [30] is an 8-unitary decoupling sequence where, following the notation above, $A_1 = A_3 = A_6 = A_8 = X$ and $A_2 = A_4 = A_5 = A_7 = Y$. In the limit of perfect control and infinitesimally short (delta) pulses, the sequence refocuses noise along any direction that varies slower than the sequence is implemented. To implement a unitary gate U within the sequence, we split it into 8 parts,

$$\begin{aligned} U_1 = U_7 = XU^{\frac{1}{8}}X, \quad U_2 = U_6 = YXU^{\frac{1}{8}}XY, \\ U_3 = U_5 = XYXU^{\frac{1}{8}}XYX, \quad U_4 = U_8 = U^{\frac{1}{8}}, \end{aligned} \tag{E.20}$$

where we simplify the expression using $XYXY = YXYX = -\mathbb{1}$, and the fact that the global phase of the desired unitary is irrelevant.

E.3.2 XY4 sequence

The XY4 sequence [31] is a 4-unitary decoupling sequence where, following the notation above, $A_1 = A_3 = X$ and $A_2 = A_4 = Y$. Like the XY8 sequence, the XY4 sequence refocuses noise along any direction that varies slower than the sequence is implemented, given that the pulses are ideal and infinitesimally short. We spread a unitary gate U across the sequence by breaking it into four parts,

$$\begin{aligned} U_1 = XU^{\frac{1}{4}}X, \quad U_2 = YXU^{\frac{1}{4}}XY, \\ U_3 = XYXU^{\frac{1}{4}}XYX, \quad U_4 = U^{\frac{1}{4}}. \end{aligned} \tag{E.21}$$

E.4 Gate Sets

Gate Set 1 (GS1) is built on Physical Model 1 using the parameters in Table E.1. This gate set was built from an XY8 pulse sequence [30], with single-qubit gates being implemented within this sequence according to the XY sequence gate protocol. Each pulse piece in the sequence was found via the GRAPE algorithm [32, 33] with control constraints from Table E.1 incorporated into the algorithm. All gates are 199.2 ns long and the discretization step for cumulant simulations was chosen to be 0.1 ns.

Gate set 2 (GS2) is built on Physical Model 2 using the parameters in Table E.2. This gate set was built from an XY4 pulse sequence [31], again, with single-qubit gates being implemented within this sequence according to the XY sequence gate protocol. All pulse

Table E.1: Parameters used for Physical Model 1, Gate Set 1.

Control Parameter		Value	Noise Parameter		Value
B_0	1.5193×10^{13}	Hz	T_1	1	s
B_1	1.5193×10^{11}	Hz	Γ_{α_1}	4.804	Hz
B_2	120		Γ_{α_2}	1.519×10^{8}	Hz
$A_{\max}$	3.798×10^{8}	Hz	τ_1	10^{-2}	s
$B_{\max}$	3.0385×10^{13}	Hz	τ_2	10^{-3}	s
$\Delta A_{\max}$	0.7596×10^{18}	Hz / s	τ_3	10^{-4}	s
$\Delta B_{\max}$	1.215×10^{23}	Hz / s	Γ_{β_1}	1.519×10^{9}	Hz
$C_{\max}$	8.73568×10^{12}	Hz	Γ_{β_2}	4.804×10^{6}	Hz
δt	10^{-10}	s	Γ_{γ}	10^{3}	Hz

pieces are performed using hard pulses. All gates are 168 ns long and the discretization step for cumulant simulations was chosen to be 0.25 ns[1].

Gate set 3 (GS3) is also built on Physical Model 2, but uses the noise parameters in Table E.3 to provide variety in the resultant gate errors. No refocusing pulse sequences are used; all gates are generated from simple hard pulses. All gates are 25 ns long and the discretization step for cumulant simulations was chosen to be 0.1 ns.

For each gate set, the two-qubit CNOT gate is implemented using the identity

$$(-1)^{\frac{3}{4}} U_{\text{CNOT}} = e^{i\frac{\pi}{2}\mathbb{1}\otimes\frac{X}{2}} e^{-i\frac{\pi}{2}\mathbb{1}\otimes\frac{Y}{2}} e^{-i\pi\frac{Z\otimes Z}{4}} e^{i\frac{\pi}{2}\mathbb{1}\otimes\frac{Y}{2}} e^{i\frac{\pi}{2}\frac{Z}{2}\otimes\mathbb{1}}. \tag{E.22}$$

For GS1 and GS2, as for single-qubit gates, the gate is broken into parts that are interspersed into their respective XY sequences. In this case however, the first two single-qubit rotations are done during the first half of the XY sequence, and the last two single-qubit rotations are done during the second half of the XY sequence, in a way similar to the single-qubit gates. The two-qubit coupling operation is implemented in the middle of the XY sequence. For GS3, the CNOT gate is implemented according to the above identity, using hard pulses.

See Appendix F.1 for details on the procedure used to simulate the gates.

[1] As GS2 uses hard pulses, the cumulant simulation can be discretized more coarsely, as the pulse amplitudes and phases remain constant for longer periods of time. The same can be said for GS3, though given that the gates are so short, a smaller time step was used anyway.

Table E.2: Parameters used for Physical Model 2, Gate Set 2

Control Parameter	Value	
$A_{\max}$	$2\pi \times 10^8$	Hz
$B_{\max}$	$2\pi \times 10^9$	Hz
$C_{\max}$	$2\pi \times 10^8$	Hz
δt	10^{-9}	s

Noise Parameter	Value	
T_1	10^{-4}	s
Γ_α	3×10^4	Hz
Γ_{β_1}	3×10^4	Hz
Γ_{β_2}	$10^6/2\pi$	Hz
$\Lambda_\alpha^{(l)}$	$1/2\pi$	Hz
$\Lambda_\alpha^{(u)}$	10^9	Hz
$\Lambda_{\beta_1}^{(l)}$	$1/2\pi$	Hz
$\Lambda_{\beta_1}^{(u)}$	10^9	Hz
$\Lambda_{\beta_2}^{(l)}$	$1/2\pi$	Hz
$\Lambda_{\beta_2}^{(u)}$	10^9	Hz
Γ_γ	$1.2 \times 10^3/2\pi$	Hz
$\Lambda_\gamma^{(l)}$	$1/2\pi$	Hz
$\Lambda_\gamma^{(u)}$	10^9	Hz

Table E.3: Parameters used for Physical Model 2, Gate Set 3

Control Parameter	Value	
$A_{\max}$	$2\pi \times 10^8$	Hz
$B_{\max}$	$2\pi \times 10^9$	Hz
$C_{\max}$	$2\pi \times 10^8$	Hz
δt	10^{-9}	s

Noise Parameter	Value	
T_1	10^{-5}	s
Γ_α	0	Hz
Γ_{β_1}	10^4	Hz
Γ_{β_2}	10^4	Hz
$\Lambda_\alpha^{(l)}$	$1/2\pi$	Hz
$\Lambda_\alpha^{(u)}$	10^9	Hz
$\Lambda_{\beta_1}^{(l)}$	$1/2\pi$	Hz
$\Lambda_{\beta_1}^{(u)}$	10^9	Hz
$\Lambda_{\beta_2}^{(l)}$	$1/2\pi$	Hz
$\Lambda_{\beta_2}^{(u)}$	10^9	Hz
Γ_γ	$1.2 \times 10^3/2\pi$	Hz
$\Lambda_\gamma^{(l)}$	$1/2\pi$	Hz
$\Lambda_\gamma^{(u)}$	10^9	Hz

Appendix F

Methods

Note, the content of this Appendix is reproduced from Appendix A of [12]. It should be considered technically separate from this thesis, and is only included for completeness.

F.1 Cumulant Simulation

To simulate quantum logic gates using our noise models, a method for the simulation of stochastic quantum evolution is required. This has been considered in the context of analyzing the fidelity with which decoherence-free subsystems can be implemented [48]. In that case, the cumulant expansion [49, 50, 51] was applied to model the effects of stochastic dynamics on a quantum system.

Following that approach, we will consider that, conditioned on a particular realization of noise, our system evolves according to the Liouville-von Neumann equation

$$\frac{\partial}{\partial t}\rho(t) = -i[H(t), \rho(t)] + D[\rho(t)], \tag{F.1}$$

where $\rho(t)$ is the density operator describing our system at time t, H is the Hamiltonian of the system, and where $D \in \mathrm{T}(\mathcal{H})$ is a linear transformation describing the decoherence of the system. We assume that $H(t)$ can be decomposed into deterministic and stochastic parts,

$$H(t) = H_{\mathrm{det}}(t) + H_{\mathrm{st}}(t). \tag{F.2}$$

We then further decompose $H_{\text{st}}(t)$ such that all of the stochasticity is encapsulated in a set of scalar-valued functions $\{\omega_1(t), \ldots, \omega_k(t)\}$. Thus,

$$H_{\text{st}}(t) = \sum_i \omega_i(t) A_i(t) \tag{F.3}$$

for some set of deterministic operator-valued functions $\{A_i(t)\}$.

To analyze the dissipation transformation D, we assume that it can be written in Lindblad form,

$$D[\rho(t)] = \sum_i L_i \rho(t) L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho(t)\}, \tag{F.4}$$

where $\{L_i\}$ are called the Lindblad operators of the system.

Both the Liouvillian operator $\mathcal{L}[\rho(t)] \equiv [H, \rho(t)]$ and the dissipation operator D act linearly on density operators, and thus may be represented by superoperators $\hat{\mathcal{L}}, \hat{D} \in \mathrm{L}(\mathrm{L}(\mathcal{H}))$, where $\mathrm{L}(\mathcal{H})$ marks the set of all linear operators acting on Hilbert space $\mathcal{H}$. Using the isomorphism that $\mathrm{L}(\mathcal{H}) \cong \mathcal{H} \otimes \mathcal{H}$, we shall use the column-stacking basis for $\mathcal{H} \otimes \mathcal{H}$, such that $||i\rangle\langle j|\rangle\rangle = |j\rangle \otimes |i\rangle$. Therefore one can rewrite Equation (F.1) as

$$\frac{\partial}{\partial t} |\rho(t)\rangle\rangle = \left(-i[\hat{\mathcal{L}}_{\text{det}}(t) + \hat{\mathcal{L}}_{\text{st}}(t)] + \hat{D}\right) |\rho(t)\rangle\rangle. \tag{F.5}$$

Now we go to the rotating frame of the deterministic superoperator $\hat{\mathcal{L}}_{\text{det}}(t)$, i.e. we define a unitary $\mathcal{U}(t) = \mathrm{T}\exp\left(-i\int_0^t \hat{\mathcal{L}}_{\text{det}}(t')\mathrm{d}t'\right)$ such that

$$\frac{\partial}{\partial t} |\tilde{\rho}(t)\rangle\rangle = \left(-i\,\mathcal{U}^\dagger(t)\hat{\mathcal{L}}_{\text{st}}(t)\mathcal{U}(t) + \mathcal{U}^\dagger(t)\hat{D}\mathcal{U}(t)\right) |\tilde{\rho}(t)\rangle\rangle, \tag{F.6}$$

where $|\tilde{\rho}(t)\rangle\rangle = \mathcal{U}^\dagger(t)\,|\rho(t)\rangle\rangle$.

The formal solution to Equation (F.6), then, for a single realization of the trajectories $\{\omega(t)\}$ is given by

$$|\tilde{\rho}(t)\rangle\rangle = \mathrm{T}\exp\left(-i\int_0^t \hat{G}(t')\mathrm{d}t'\right) |\rho(0)\rangle\rangle, \tag{F.7}$$

with $\hat{G}(t) \equiv \mathcal{U}^\dagger(t)\hat{\mathcal{L}}_{\text{st}}(t)\mathcal{U}(t) + i\,\mathcal{U}^\dagger(t)\hat{D}\mathcal{U}(t)$.

For our purposes, we are interested in the average evolution $\hat{S}$ over the ensemble of control trajectories,

$$\hat{S}(t) = \left\langle \mathrm{T}\exp\left(-i\int_0^t \hat{G}(t')\mathrm{d}t'\right)\right\rangle. \tag{F.8}$$

The cumulant expansion gives us that $\hat{S}(t) = \exp(\hat{K}(t))$, where

$$\hat{K}(t) = \sum_{n=1}^{\infty} \frac{(-it)^n}{n!} K_n = -it\hat{K}_1 - \frac{t^2}{2}\hat{K}_2 + \cdots, \tag{F.9}$$

$$\hat{K}_1 = \frac{1}{t}\int_0^t \mathrm{d}t_1 \left\langle \hat{G}(t_1) \right\rangle, \tag{F.10}$$

$$\hat{K}_2 = \frac{1}{t^2}\mathrm{T}\int_0^t \mathrm{d}t_1 \int_0^t \mathrm{d}t_2 \left\langle \hat{G}(t_1)\hat{G}(t_2) \right\rangle - \hat{K}_1^2. \tag{F.11}$$

To simplify this, we assume that each control parameter $\omega_i(t)$ is a trajectory of a stationary zero-mean process (The zero-mean assumption technically isn't an assumption; the mean of each random process can be absorbed into the deterministic part of the Hamiltonian.) That is, that $\vec{\omega} \sim \mathrm{GP}(0, \vec{\vec{\phi}})$, where $\vec{\vec{\phi}}$ is the matrix-valued autocorrelation function for $\vec{\omega}(t)$, such that $\phi_{i,j}(t_1 - t_2) = \langle \omega_i(t_1)\omega_j(t_2)\rangle$.

Then $\hat{K}_1$ becomes simply

$$\hat{K}_1 = \frac{i}{t}\int_0^t \mathrm{d}t_1 \, \mathcal{U}^\dagger(t_1)\hat{D}\mathcal{U}(t_1), \tag{F.12}$$

whereas we can then rewrite $\hat{K}_2$ in terms of the autocorrelation function,

$$\hat{K}_2 = \frac{2}{t^2}\int_0^t \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2 \sum_{i,j=1}^{k} \phi_{i,j}(t_1 - t_2)\mathcal{U}^\dagger(t_1)\hat{A}_i(t_1)\mathcal{U}(t_1)\mathcal{U}^\dagger(t_2)\hat{A}_j(t_2)\mathcal{U}(t_2) \tag{F.13}$$

$$- \frac{2}{t^2}\int_0^t \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2 \, \mathcal{U}^\dagger(t_1)\hat{D}\mathcal{U}(t_1)\mathcal{U}^\dagger(t_2)\hat{D}\mathcal{U}(t_2) - \hat{K}_1^2, \tag{F.14}$$

where $\hat{A}_i(t) = -A_i^*(t) \otimes \mathbb{1} + \mathbb{1} \otimes A_i(t)$. In this way, we note that the cumulant expansion generalizes the Magnus expansion so as to account for stochastically-varying fields. The motivation for using cumulants instead of expanding the time-ordered exponential in terms of moments of the stochastic process stems from the fact that cumulant averages enter in the exponential, reducing the risk of truncation artefacts.

To numerically simulate the gate action, we discretize $\hat{\mathcal{L}}_{\mathrm{det}}(t')$ along our gate length t at N points, with equal time intervals Δt between these points, i.e. we evaluate $\{\hat{\mathcal{L}}_{\mathrm{det}}(m\Delta t)\}$, with $m = 1, ..., N$ while $t = N\Delta t$. Next we approximate $\mathcal{U}(n\Delta)$ by

$$\mathcal{U}(n\Delta) \approx \exp\left(-i\hat{\mathcal{L}}_{\mathrm{det}}(n\Delta t)\Delta t\right) \; ... \; \exp\left(-i\hat{\mathcal{L}}_{\mathrm{det}}(\Delta t)\Delta t\right)\exp\left(-i\hat{\mathcal{L}}_{\mathrm{det}}(0)\Delta t\right). \tag{F.15}$$

Finally, we turn turn the integral in line (F.12) into a sum

$$\hat{K}_1 \approx \frac{i}{N} \sum_{n=0}^{N-1} \mathcal{U}^\dagger(n\Delta t)\hat{D}\mathcal{U}(n\Delta t), \tag{F.16}$$

and the double integral in line (F.13) into a double sum

$$\begin{aligned}
\hat{K}_2 \approx & \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{i,j=1}^{k} \phi_{i,j}(0)\mathcal{U}^\dagger(n\Delta t)\hat{A}_i(n\Delta t)\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(n\Delta t)\hat{A}_j(n\Delta t)\mathcal{U}(n\Delta t) \\
& + \frac{2}{N^2} \sum_{n=1}^{N-1} \sum_{m=0}^{n-1} \sum_{i,j=1}^{k} \phi_{i,j}((n-m)\Delta t)\mathcal{U}^\dagger(n\Delta t)\hat{A}_i(n\Delta t)\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(m\Delta t)\hat{A}_j(m\Delta t)\mathcal{U}(m\Delta t) \\
& - \frac{1}{N^2} \sum_{n=0}^{N-1} \mathcal{U}^\dagger(n\Delta t)\hat{D}\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(n\Delta t)\hat{D}\mathcal{U}(n\Delta t) \\
& - \frac{2}{N^2} \sum_{n=1}^{N-1} \sum_{m=0}^{n-1} \mathcal{U}^\dagger(n\Delta t)\hat{D}\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(m\Delta t)\hat{D}\mathcal{U}(m\Delta t) - \hat{K}_1^2.
\end{aligned} \tag{F.17}$$

To simulate the gates, we truncated $\hat{K}(t)$ in Equation (F.9) at second order, which can be partially justified with the following. If we have no dissipator term in Equation (F.1), then due to statistical independence, the mth order cumulant disappears if, for a set of times $\{t_1, t_2, ..., t_n\}$, any of the time gaps $|t_1 - t_2|$, $|t_2 - t_3|$, ...,$|t_{n-1} - t_n|$ are larger than the correlation time τ_c of the stochastic process [49]. Since cumulants at every order vanish once the gap between the set of time points exceeds τ_c, then if $t \gg \tau_c$, the mth order cumulant $\hat{K}_m$ is effectively an integral over an $(m-1)$ dimensional sphere with radius τ_c, integrated over t. Therefore, $\hat{K}_m$ scales roughly as $\tau_c^{m-1}A^m t$, where A is the maximum norm of $\hat{A}_i(t)$. Comparing the second- and fourth-order cumulants, $\hat{K}_2$ and $\hat{K}_4$, reveals that $\frac{\tau_c^3 A^4 t}{\tau_c A^2 t} = \tau_c^2 A^2$, meaning that if $\tau_c A \ll 1$ and $t \gg \tau_c$, we have a justification for truncating the cumulant expansion at second order. For the physical models considered in this work, the dissipator terms in the Liouville-von Neumann equation were considerably smaller in their norm than the noise Hamiltonian terms, so we assume that the arguments above are still applicable.

F.2 Circuit Simulation

Each gate in the gate set acts on either one or two qubits, subjecting the rest to identical, uncorrelated noise (the noisy identity gate). The action of a noisy process on a quantum register can be calculated in the Kraus representation,

$$\Lambda(\rho) = \sum_{j=1}^{4^n} A_j \rho A_j^\dagger. \tag{F.18}$$

For a generic noisy process, using naïve matrix multiplication, this calculation involves $\sim 2^{5n}$ operations and requires the storage of $\sim 2^{4n}$ complex parameters. These costs can be reduced dramatically by exploiting the fact that the noise is independent and that the gate set acts identically on different qubits. Noise maps that act independently commute, and can be applied in sequence as a result:

$$\lambda^{\otimes n} = \lambda_1(\lambda_2(\ldots\lambda_n(\rho))) \tag{F.19}$$

Thus, the amount of storage is reduced to that required to store the gate set and the current state, $\sim 2^{2n}$ complex parameters, and the number of operations required now scales as $n2^{3n}$. This can be further reduced by noting that each channel λ_j is equivalent to the perfect identity on $n-1$ qubits, and its effect can be pre-calculated to reduce the total number of operations to $n2^{2n}$. The extension to two-qubit gates is straightforward; for further information, see [52].

References

[1] Austin G. Fowler, Ashley M. Stephens, and Peter Groszkowski. High threshold universal quantum computation on the surface code. *arXiv:0803.0272*, March 2008. Phys. Rev. A 80, 052312 (2009).

[2] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, October 1997.

[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*, STOC '96, page 212–219, New York, NY, USA, 1996. ACM.

[4] Nathan Wiebe, Daniel Braun, and Seth Lloyd. Quantum algorithm for data fitting. *Physical Review Letters*, 109(5):050505, August 2012.

[5] Panos Aliferis, Daniel Gottesman, and John Preskill. Accuracy threshold for post-selected quantum computation. *arXiv:quant-ph/0703264*, March 2007. Quant. Inf. Comput. 8 (2008) 181-244.

[6] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, February 2005.

[7] Sergey Bravyi, Guillaume Duclos-Cianci, David Poulin, and Martin Suchara. Subsystem surface codes with three-qubit check operators. *arXiv:1207.1443*, July 2012.

[8] David S. Wang, Austin G. Fowler, Ashley M. Stephens, and Lloyd Christopher L. Hollenberg. Threshold error rates for the toric and planar codes. *Quantum Information & Computation*, 10(5&6):456–469, 2010.

[9] Austin G. Fowler. Proof of finite surface code threshold for matching. *Physical Review Letters*, 109(18):180502, November 2012.

[10] Sergey Bravyi and Alexander Vargo. Simulation of rare events in quantum error correction. arXiv e-print 1308.6270, August 2013.

[11] Easwar Magesan, Daniel Puzzuoli, Christopher E. Granade, and David G. Cory. Modeling quantum noise for efficient testing of fault-tolerant circuits. *Physical Review A*, 87(1):012324, January 2013.

[12] Daniel Puzzuoli, Christopher Granade, Holger Haas, Ben Criger, Easwar Magesan, and D. G. Cory. Tractable simulation of error correction with honest approximations to realistic fault models. *Physical Review A*, 89(2):022306, February 2014.

[13] John Watrous. Theory of quantum information course notes. 2011.

[14] A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, April 1976.

[15] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216 –1227, May 1999.

[16] Isaac L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997.

[17] Michael A Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *quant-ph/0205035*, May 2002. Phys. Lett. A 303 (4): 249-252 (2002).

[18] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson. Scalable protocol for identification of correctable codes. *Physical Review A*, 78(1):012347, July 2008.

[19] C.W. Helstrom. *Quantum Detection and Estimation Theory*. Mathematics in Science and Engineering : a series of monographs and textbooks. Academic Press, 1976.

[20] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, 2002.

[21] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, November 2004.

[22] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson. Negative quasi-probability as a resource for quantum computation. *New Journal of Physics*, 14(11):113011, November 2012.

[23] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, April 2002.

[24] M. Van den Nest. Efficient classical simulations of quantum fourier transforms and normalizer circuits over abelian groups. arXiv e-print 1201.4867, January 2012.

[25] Juan Bermejo-Vega and Maarten Van den Nest. Classical simulations of abelian-group normalizer circuits with intermediate measurements. arXiv e-print 1210.3637, October 2012.

[26] Richard Jozsa and Maarten Van den Nest. Classical simulation complexity of extended Clifford circuits. *arXiv:1305.6190*, May 2013.

[27] John Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217–238, 2009.

[28] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 1.22. http://cvxr.com/cvx, September 2012.

[29] Mauricio Gutiérrez, Lukas Svec, Alexander Vargo, and Kenneth R. Brown. Approximation of realistic errors by Clifford channels and Pauli measurements. *Physical Review A*, 87(3):030302, March 2013.

[30] Terry Gullion, David B Baker, and Mark S Conradi. New, compensated Carr-Purcell sequences. *Journal of Magnetic Resonance (1969)*, 89(3):479–484, October 1990.

[31] A.A Maudsley. Modified Carr-Purcell-Meiboom-Gill sequence for NMR Fourier imaging applications. *Journal of Magnetic Resonance (1969)*, 69(3):488–491, October 1986.

[32] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J. Glaser. Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *Journal of Magnetic Resonance*, 172(2):296–305, February 2005.

[33] Troy W. Borneman, Martin D. Hürlimann, and David G. Cory. Application of optimal control to CPMG refocusing pulse design. *Journal of Magnetic Resonance*, 207(2):220–233, December 2010.

[34] E. Knill. Quantum computing with realistically noisy devices. *Nature*, 434(7029):39–44, Mar 2005.

[35] Christopher Granade and Ben Criger. QuaEC: Quantum error correction analysis in python, 2012.

[36] Michael R. Geller and Zhongyuan Zhou. Efficient error models for fault-tolerant architectures and the pauli twirling approximation. *Physical Review A*, 88(1):012314, July 2013.

[37] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, 1 edition, January 2000.

[38] C. King and M.B. Ruskai. Minimal entropy of states emerging from noisy quantum channels. *Information Theory, IEEE Transactions on*, 47(1):192–209, 2001.

[39] M. F. Sacchi. Minimum error discrimination of Pauli channels. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S333, October 2005.

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

[41] J. Schlienz and G. Mahler. Description of entanglement. *Physical Review A*, 52(6):4396–4404, December 1995.

[42] G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics (1965-1997)*, 48(2):119–130, 1976.

[43] Vittorio Gorini, Andrzej Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17(5):821–825, May 1976.

[44] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, USA, August 2002.

[45] L. D. Landau and E. M. Lifshitz. *Statistical physics. Pt.1*. 1969.

[46] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[47] F.N. Hooge and P.A. Bobbert. On the correlation function of $1/f$ noise. *Physica B: Condensed Matter*, 239(3–4):223–230, August 1997.

[48] P. Cappellaro, J. S Hodges, T. F Havel, and D. G Cory. Principles of control for decoherence-free subsystems. *The Journal of Chemical Physics*, 125(4):044514–044514–10, July 2006.

[49] Ryogo Kubo. Generalized cumulant expansion method. *Journal of the Physical Society of Japan*, 17(7):1100–1120, July 1962.

[50] N. G. van Kampen. A cumulant expansion for stochastic linear differential equations. I. *Physica*, 74:215–238, June 1974.

[51] N. G. van Kampen. A cumulant expansion for stochastic linear differential equations. II. *Physica*, 74:239–247, June 1974.

[52] J.R. Johansson, P.D. Nation, and Franco Nori. QuTiP 2: A Python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 184(4):1234 – 1240, 2013.

[53] M. H. Devoret and R. J. Schoelkopf. Superconducting circuits for quantum information: An outlook. *Science*, 339(6124):1169–1174, March 2013. PMID: 23471399.

[54] 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, September 2004.

[55] Troy W. Borneman. Ph. D. Thesis. techniques for noise suppression and robust control in spin-based quantum information processors. *Massachusetts Institute of Technology*, December 2012.

[56] Daniel A. Lidar and K. Birgitta Whaley. Decoherence-free subspaces and subsystems. *arXiv:quant-ph/0301032*, January 2003. in "Irreversible Quantum Dynamics", F. Benatti and R. Floreanini (Eds.), pp. 83-120 (Springer Lecture Notes in Physics vol. 622, Berlin, 2003).

[57] D. A. Lidar, D. Bacon, and K. B. Whaley. Concatenating decoherence-free subspaces with quantum error correcting codes. *Physical Review Letters*, 82(22):4556–4559, May 1999.

[58] D. A. Lidar, I. L. Chuang, and K. B. Whaley. Decoherence-free subspaces for quantum computation. *Physical Review Letters*, 81(12):2594, 1998.

[59] Daniel Gottesman. An introduction to quantum error correction and fault-tolerant quantum computation. *arXiv:0904.2557*, April 2009.

[60] Ben W. Reichardt. Error-detection-based quantum fault tolerance against discrete Pauli noise. *arXiv:quant-ph/0612004*, November 2006.

[61] Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill. Topological quantum memory. *arXiv:quant-ph/0110143*, October 2001. J. Math. Phys. 43, 4452-4505 (2002).

[62] N. Cody Jones, Rodney Van Meter, Austin G. Fowler, Peter L. McMahon, Jungsang Kim, Thaddeus D. Ladd, and Yoshihisa Yamamoto. Layered architecture for quantum computing. *Physical Review X*, 2(3):031007, July 2012.

[63] Daniel Gottesman. Stabilizer codes and quantum error correction. *quant-ph/9705052*, May 1997.

[64] Panos Aliferis, Daniel Gottesman, and John Preskill. Quantum accuracy threshold for concatenated distance-3 codes. *arXiv:quant-ph/0504218*, April 2005. Quant. Inf. Comput. 6 (2006) 97-165.

[65] Cecilia C. López, Ariel Bendersky, Juan Pablo Paz, and David G. Cory. Progress toward scalable tomography of quantum maps using twirling-based methods and information hierarchies. *Physical Review A*, 81(6):062113, June 2010.

[66] 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, September 2007. PMID: 17901327.

[67] F. T. Hioe and J. H. Eberly. n-level coherence vector and higher conservation laws in quantum optics and quantum mechanics. *Physical Review Letters*, 47(12):838–841, September 1981.

[68] Yu Tomita, Mauricio Gutiérrez, Chingiz Kabytayev, Kenneth R. Brown, M. R. Hutsel, A. P. Morris, Kelly E. Stevens, and G. Mohler. Comparison of ancilla preparation and measurement procedures for the steane [[7,1,3]] code on a model ion trap quantum computer. *arXiv:1305.0349*, May 2013.

[69] Christopher E Granade, Christopher Ferrie, Nathan Wiebe, and D G Cory. Robust online Hamiltonian learning. *New Journal of Physics*, 14(10):103013, October 2012.

[70] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and D. G. Cory. Hamiltonian learning and certification using quantum resources. (1309.0876), September 2013.