Characterizing Noise in Quantum Systems

by

Easwar Magesan

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

Waterloo, Ontario, Canada, 2012

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

In practice, quantum systems are not completely isolated from their environment and the resulting system-environment interaction can lead to information leakage from the system. As a result, if a quantum system is to be used for storing or manipulating information, one would like to characterize these environmental noise effects. Such a characterization affords one the ability to design robust methods for preserving the information contained in the system. Unfortunately, completely characterizing the noise in a realistic amount of time is impossible for even moderately large systems.

In this thesis we discuss methods and diagnostics for partially characterizing quantum noise processes that are especially useful in quantum information and computation. We present a randomized benchmarking protocol that provides a scalable method for determining important properties of the noise affecting the set of gates used on a quantum information processor. We also prove various properties of the quantum gate fidelity, which is a useful state-dependent measure of the distance between two quantum operations as well as an important diagnostic of the noise affecting a quantum process. Some non-intuitive generic features of quantum operations acting on large-dimensional quantum systems are also presented.

Acknowledgements

I would like to thank my supervisor Joseph Emerson, David Cory, Jay Gambetta, and my committee members Achim Kempf and Ray Laflamme, for the help and guidance they have provided me throughout my graduate studies at Waterloo. I would also like to thank Kevin Resch and Patrick Hayden for the time and effort they have put into being examiners for my defence. I am indebted to the people I have collaborated with throughout my graduate studies, in particular, Troy Borneman, David Cory, Joseph Emerson, Robin Blume-Kohout, Jay Gambetta, David Kribs, Dan Puzzuoli, and Marcus Silva. My research would have been much more difficult without their insight and creativity.

I have been fortunate to have discussions with many great researchers during my time at Waterloo and I would like to thank all of them. This lengthy list includes, but is certainly not limited to, Richard Cleve, Chris Ferrie, Daniel Gottesman, Chris Granade, Ian Hincks, Martin Laforest, Iman Marvian, Ryan Morris, Osama Moussa, Yingkai Ouyang, Gina Passante, Marco Piani, Joseph Rebstock, Cozmin Ududec, Victor Veitsch and John Watrous. I have had fantastic instructors as a graduate student and I thank every one of them for their patience and enthusiasm.

I am deeply indebted to the University of Waterloo, the Institute for Quantum Computing and the Perimeter Institute for the resources they have provided me during my time at Waterloo. I also thank Raymond Laflamme and Michele Mosca for the work they have put in to making IQC such a wonderful institute. It is amazing how much IQC has grown since I first arrived as a graduate student. I am indebted to the University of Waterloo, the Institute for Quantum Computing, NSERC, CIFAR and the government of Ontario for the funding they have provided me throughout my graduate studies.

I thank all of the administrative support staff in the applied math department and at IQC; their assistance during my time at Waterloo is greatly appreciated. Helen Warren has been an immense help in keeping me organized and ensuring my studies went smoothly. Her kindness, patience and guidance are deeply appreciated. I would also like to thank Wendy Reibel for all of the help she has provided me during my graduate studies.

Finally, I thank my family and friends for their support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Quantum information theory is the study of representing and transforming information using the principles of quantum mechanics. The information is encoded into the set of states of the quantum system and transformed via operations that are quantum mechanical in nature. Thinking about information from a quantum perspective has led to a variety of interesting and novel results. Some examples of which are quantum algorithms that can solve problems faster than any current known classical algorithm, with the speed-up being exponential in certain cases. Specific examples include Shor's factorization algorithm [131], Grover's search algorithm [60], simulating the evolution of physical systems [94] and solving certain linear systems of equations [62]. Quantum information theory has also provided the first provably secure key distribution scheme [13], called the "BB84" protocol, and has led to a deeper understanding of computational complexity and the relationship between various complexity classes [72, 143]. The field of quantum complexity theory allows for a better determination of the computational problems that can be solved efficiently under specific resource assumptions.

The advantages of using quantum theory for information processing can only be realized if the physical implementation of a large-scale quantum information processor is possible. The explicit form of the operations used to process the information depends on the particular model of computation being used, examples of which are the standard circuit [43], measurement-based [58, 120], adiabatic [49], and topological models of computation [77]. These are all equivalent in that any computation performed using one model can be simulated in any of the other models. Hence these models are all capable of performing universal quantum computation. In many of these models, including the standard circuit model, the ability of the "quantum computer" to perform true quantum computation can be measured against certain criteria, called the DiVincenzo criteria [45]:

- scalability of the physical system in the number of well-defined subsystems (qubits),
- the ability to efficiently initialize qubits to a standard input state,
- the time for which the system remains quantum is much longer than the gate-operation time,
- the ability to perform a universal set of gates,
- the ability to perform qubit-specific measurements.

Two additional criteria that allow for any quantum communication protocol were given in [45] and, for completeness, these are:

- the ability to interconvert between stationary and flying qubits,
- the ability to faithfully transmit flying qubits between locations.

There are various proposals for the implementation of a quantum information processor, for instance using NMR [37, 55, 36], ion traps [34], superconducting circuits [18], NV centres [61], optical lattices [22] and quantum dots [96]. To date, implementations have been rudimentary from the perspective of the ultimate goal and there is debate as to whether large scale quantum information processors will be a reality in the future. The reasons for the difficulty in constructing a large-scale implementation vary across the different implementation schemes. Perhaps the most important difficulty is the extreme sensitivity of quantum systems to their environment. More precisely, quantum systems tend to lose their quantum "coherent" nature (decohere) on time scales much faster than those needed to perform complex computations (DiVincenzo's third criterion). As a result an important area of research is how to engineer quantum systems such that environmental effects are minimized.

It is interesting to note that it may not always be the case that environmental effects will be detrimental to observing quantum effects on long time-scales. For instance there is recent experimental evidence that quantum effects play an important role in the energy transport mechanism of photosynthesis in certain plant species [48, 35]. The very possibility

that quantum coherence plays a significant role is surprising since the system involved in the transport is not well isolated from its environment and thus intuitively should decohere on time-scales much faster than those for which transport occurs. One explanation that has been put forth is that the the geometric arrangement of the molecules making up the system relative to the environment has been optimized over time so that the environment actually assists in the coherent transport of energy (a phenomenon called "noise-assisted transport") [115].

The fact that environmental interactions can significantly alter the state of a quantum system leads to the question of whether one can devise clever methods for hiding information in the quantum system so that the information can not leak out to the environoment. If this were possible then the information would be preserved throughout the environmental interaction and in principle could be recovered at the end of the process. The idea that certain states can be preserved, or corrected for, under environmental noise interactions forms the basis of "quantum error-correction" [132, 27, 135, 82], which is of fundamental importance for the experimental realization of quantum computation. Unlike classical error models which primarily deal with bit-flip and erasure errors, there is a large variety of possible errors affecting a quantum system. As well if one is to correct for an error, a measurement must be performed on the system. Such a measurement can alter the state of the system thus potentially destroying the information regarding what error occurred. These nuances, among other subtleties, make it surprising that quantum error-correction is even possible in the first place. Fortunately quantum error-correction is possible and various error-correction methods have been devised to combat noise effects.

Analogous to classical error-correction leading to a threshold theorem for fault-tolerant classical computation, the field of quantum error-correction has enabled the development of a threshold theorem for fault-tolerant quantum computation [2, 83, 118]. The main idea behind the threshold theorem is that for certain noise models there exist fault-tolerant encoding schemes such that arbitrarily precise fault-tolerant quantum computation is possible provided the error rate on the physical gates is below a certain threshold value. Sec.'s B.3 and B.4 contain further discussion on quantum error correction and fault-tolerance.

From the above discussion, it is clear that a detailed understanding of the noise affecting a quantum system is desirable for the design of good error-correcting codes. Indeed, it is often the case that a large savings in the overhead required for error-correction and fault-tolerance protocols can be obtained when one does not have to correct for arbitrary noise models. Unfortunately, the number of parameters required to completely describe a noise process grows exponentially in the number of subsystems comprising the system [33, 117]. Hence, just obtaining a complete description of the noise process via a method such as quantum process tomography (QPT) [33] is not a scalable process.

Since, in general, a complete characterization of the noise affecting a quantum system will not be available, an important task is to find efficient methods for *partially characterizing* the noise process. This idea constitutes the first of two main research areas of this thesis and is contained in Chapter 2. We present a scheme called *randomized benchmarking* that provides a characterization of the performance of the set of gates implemented by a quantum information processor via a single parameter. The scheme is ideal for benchmarking quantum gates in that it is both scalable as well as robust against state-preparation and measurement errors (except in extreme cases). The articles pertaining to this research area are given by [99, 100].

The second area of research is contained in Chapter 3 and is based on the articles [97, 98]. We discuss and prove various properties of a mathematical quantity called the *quantum gate fidelity* which is an experimentally useful measure for characterizing how far apart an intended quantum gate is from the operation that is actually implemented. In particular, the randomized benchmarking protocol of Chapter 2 utilizes the average of the gate fidelity over all input states to characterize the performance of a quantum information processor.

The results I present in Chapter's 2 and 3 constitute the parts of the research for which I was a contributing member. The vast majority of the material required for understanding the results of the thesis is contained in the appendix and is referenced when necessary. The reason for this is to make the presentation as smooth as possible by not having to introduce new concepts throughout the presentation. Also, the appendix starts from the very basic concepts of quantum mechanics and quantum information theory. This choice was made to assist readers unfamiliar with certain aspects of the subject as well as to make the presentation as self-contained as possible. I hope the choices I have made allow for a more enjoyable reading experience.

Chapter 2

Randomized Benchmarking of Quantum Gates

One of the main challenges in building a quantum information processor is that complete noise characterization via quantum process tomography (QPT) is not scalable in the number of subsystems (qubits), n, comprising the system [33, 117]. Complete characterization of the noise is ideal because it allows for the determination of good error-correction schemes and the verification of assumptions used in fault-tolerance, such as estimates of the threshold value. There have been various methods proposed as alternatives to QPT such as ancilla-assisted process tomography [7], direct characterization of quantum dynamics (DCQD) [102] and compressive sensing methods [129]. Ancilla-assisted process tomography maps the problem to one of performing state tomography on a larger system by utilizing the Choi-Jamiolkowski isomorphism between quantum operations and quantum states on a larger system (see Sec. A.2). DCQD takes this one step further by eliminating the overhead required for performing quantum state estimation from the experimental data. Compressive sensing methods are based on the idea that when the ideal operation is unitary it can be described by a maximally sparse "process matrix" (see Sec. A.2) in a basis consisting of the intended unitary process. Hence it is expected that a reasonably precise implementation of the intended process will be described by a sparse matrix in this basis. Classical compressive sensing is then utilized to efficiently characterize the sparse matrix and thus the implemented process.

While these methods have their advantages in certain situations, the general non-scalability of QPT implies that for even moderately large quantum systems one can only partially characterize the process at hand. Various methods have been proposed for the

efficient partial characterization of processes, such as symmetrization of quantum operations [46, 47, 134, 104], selective partial tomography [11, 126] and Monte-Carlo based methods [38, 50]. Symmetrization is based on the notion of twirling quantum processes (see Sec. C.2) under a subset of the full unitary group $U(d)$. When the twirling set reflects a particular symmetry within $U(d)$, the result of the twirl is a quantum operation that is invariant with respect to this symmetry. Twirling over certain subgroups provide an exponential reduction in the number of parameters describing the process and ideally these parameters can be efficiently estimated via experiments.

Selective partial tomographic methods on the other hand provide the ability to estimate any element of the χ-matrix (see Sec. A.2) by utilizing the fact that any such element can be identified with the average fidelity (see Sec. A.4.2) of a different quantum operation. Hence implementing a quantum circuit that effectively performs this different quantum operation and estimating the average fidelity gives the desired χ-matrix element. Monte-Carlo based methods arise from first looking at estimating the fidelity between two quantum states (see Sec. A.4.1) by writing the states in terms of an orthonormal and Hermitian operator basis (when $d = 2^n$ the normalized Pauli operators form such a basis; see Sec. B.1.1). The case of particular interest is when we have a "target" state and an implemented state that ideally matches the target perfectly. One can then write an expression for the fidelity in terms of a probability distribution over the coefficients of the target state. When the target state is a stabilizer state [57], efficient sampling of this distribution can be performed using Monte-Carlo methods. The problem of estimating the average fidelity of a quantum operation can be analyzed from this protocol via utilizing the aforementioned Choi-Jamiolkowski isomorphism between quantum states and operations.

One important point regarding all of these methods for obtaining tomographic data is that each suffers from certain drawbacks in terms of either assumptions on available resources or the form of the noise. Some specific examples are:

- state-preparation and measurement errors are negligible,
- high fidelity local operations are available,
- ancilla states are available and entangled states can be prepared with high fidelity,
- classical-post processing of data can be performed efficiently to high precision,

- the noisy process is close to the intended operation.

In the context of quantum computation, it is desirable to have an explicit method for benchmarking the set of quantum gates used on a quantum information processor. Ideally the method will suffer from as few of the above drawbacks as and it will be independent of the particular implementation on which the processor has been realized. Since the complete set of quantum gates on an n qubit system is given by the unitary group $U(d)$ ($d = 2^n$) there are various problems with obtaining a benchmark for the complete set of gates, perhaps the most important of which is that $U(d)$ is a continuous group with a number of independent parameters that scales exponentially in n. Hence just generating an arbitrary element from the group is exponentially hard in n.

As a result, it would be useful to benchmark a discrete set of gates such that one obtains a reasonable indication of the reliability of the full gate set represented by $U(d)$. Such a set of gates is given by the Clifford group on n qubits, denoted Clif_n (see Sec. B.2). One reason for why Clif_n is useful in this context is that $U(d)$ can be generated by adding just one additional single-qubit gate not in the Clifford group (such as the $\frac{\pi}{8}$-gate). Moreover, the $\frac{\pi}{8}$-gate can be implemented using an ancilla magic state, Clifford operations, and a measurement in the computational basis. Thus, there exists a model of universal quantum computation such that the only gates which need to be applied are Clifford gates. In this case, a benchmark for Clif_n contains important information for the performance of a quantum information processor. This idea translates well to the fault-tolerant setting since most encoding schemes for fault-tolerant quantum computing are based on stabilizer codes. For many stabilizer codes, encoded Clifford operations are comprised of "0-level" (physical) Clifford operations, a famous example of which is the seven-qubit Steane code where Clifford operations can be applied transversally. As well, for stabilizer codes, the encoding gate can always be taken to be a Clifford operation. Thus, one expects that fault-tolerant quantum computation will be dominated by Clifford gates. Putting all of this together it is evident that benchmarking Clif_n provides significant information regarding the performance of the full set of gates used on a quantum information processor.

One proposal for obtaining a benchmark of the Clifford group is given by *randomized benchmarking* [46, 84]. The randomized benchmarking protocol in Ref. [84] outlines an experimental method for estimating the average error-rate for single qubit Clifford gates. The simplicity of this protocol has motivated experimental implementations in atomic ions for different types of traps [84, 15], NMR [124], superconducting qubits [32, 31], and atoms in optical lattices [110].

In the protocol of [84] one fits the observed fidelity decay averaged over sequences of random Clifford and Pauli gates to an exponential (in the sequence length) and *assumes*

that the decay rate gives an estimate of the average error probability per Clifford gate [46, 92, 84, 40]. However, it is unknown exactly when this assumption of an exponential decay is valid, specifically in the realistic case of gate-dependent and time-dependent noise. Moreover, extensions of the protocol to multi-qubit systems are not well understood, and scalability of the protocol is unknown. It is also easy to construct examples where the decay rate estimated via the RB protocols of Refs. [46, 84] is not reliable. An extreme but intuitively simple example is when the error is gate-dependent and equal to the exact inverse of the target gate. The error rate given by the protocol is always equal to 0 however there is substantial error on each gate.

Recently, we proposed a scalable and robust multi-qubit randomized benchmarking protocol for Clifford gates which overcomes these shortcomings [99, 100]. This chapter will be devoted to providing a detailed analysis of our results. The protocol is provably valid under the assumption of weak time and gate-dependence which improves upon the the restriction of time and gate-independent noise in [46]. We prove that for time-independent and gate-independent errors the fidelity decay is indeed given by a zeroth order fitting model which is exponential at a rate that determines the average error-rate of the noise affecting the gate set.

Using a perturbative argument we derive a first-order fitting model for the observed fidelity decay which includes correction terms due to time and gate-dependence in the errors. This formula shows that weak time and gate-dependence in the errors can lead to a deviation from the exponential decay (defining a partial test for such effects in the noise), which is illustrated via numerical examples. Moreover, the fitting models account for state preparation and measurement errors except in extreme cases since they show up as independent fit parameters in the formula. We provide a detailed proof that our protocol requires at most $O(n^2)$ quantum gates, $O(n^4)$ cost in classical pre-processing (to select each gate-sequence), and a number of single-shot repetitions that is independent of n. In the case of Pauli errors we give some novel preliminary results regarding the relationship between the benchmarking average error rate and the more common diamond norm error measure [3, 79] used in the theory of fault-tolerance.

The chapter is structured as follows: In Sec. 2.1.1 we discuss the proposed protocol in complete detail (Sec. F.1 contains the experimental protocol for implementing randomized benchmarking). Sec. 2.1.2 provides a detailed analysis of the perturbative expansion, as well as expressions for the zeroth and first order fitting models. Sec. 2.2 provides a sufficient condition for neglecting higher order terms in the model as well as a simple case for when the benchmarking scheme fails. We also discuss when the protocol is robust against state preparation and measurement errors. Sec. 2.5 discusses the relationship between the error rate given by the benchmarking scheme and other measures of error commonly used in

quantum information. Sec. 2.6 provides a detailed proof that our protocol is scalable in the number of qubits comprising the system and Sec. 2.7 provides numerical examples illustrating the protocol as well as an example for which the fidelity decay must be modelled using the first order model. Lastly, concluding remarks and a discussion for potential areas of further research are contained in Sec.'s 2.8 and 2.9.

2.1 Randomized Benchmarking

Loosely speaking, the main idea behind our protocol is to apply random sequences of Clifford gates to an input state and observe the fidelity decay as the sequence length increases. Before going into details, let us first set some notation and make various definitions that will be used throughout the presentation.

Denote the elements of Clif_n by $\mathcal{C}_i$, $i \in |\text{Clif}_n|$, and set M to be the maximum length of sequences consisting of Clifford gates that will be used in the randomized benchmarking protocol. Suppose that the actual implementation of $\mathcal{C}_i$ at time j $(1 \leq j \leq M)$ results in the map $\mathcal{E}_{i,j}$ with

$$\mathcal{E}_{i,j} = \Lambda_{i,j} \circ \mathcal{C}_i \tag{2.1}$$

for some error map $\Lambda_{i,j}$. Hence to each Clifford $\mathcal{C}_i$ we associate a sequence $(\Lambda_{i,1}, ..., \Lambda_{i,M})$ which represents the time-dependent noise operators affecting $\mathcal{C}_i$. We define the average error operator as follows,

Definition 1. *Average Error Operator*

The average error operator affecting the gates in Clif$_n$ is given by,

$$\Lambda = \frac{1}{M\,|Clif_n|} \sum_j \sum_i \Lambda_{i,j}. \tag{2.2}$$

Consider the twirl of the average error operator over Clif_n. As discussed in Sec. C.2 this produces a depolarizing channel Λ_d,

$$\begin{aligned} \Lambda_d(\rho) &= \frac{1}{|\text{Clif}_n|} \sum_i C_i^\dagger \circ \Lambda \circ C_i \,(\rho) \\ &= p\rho + (1-p)\frac{\mathbb{1}}{d} \end{aligned} \tag{2.3}$$

and the average fidelity of Λ, denoted F_{ave} (see Sec. A.4.2), is invariant under the twirl (see Sec. A.2 for a discussion of depolarizing quantum channels). Hence,

$$F_{\text{ave}} = p + \frac{1-p}{d}. \tag{2.4}$$

We define the average error rate of the set of Clifford gates as follows:

Definition 2. *Average Error Rate*

The average error rate, r, of the Clifford gates used for quantum computation is defined to be,

$$\begin{aligned} r &= 1 - F_{\text{ave}} \\ &= 1 - \left(p + \frac{1-p}{d}\right) \\ &= \frac{(d-1)(1-p)}{d}. \end{aligned} \tag{2.5}$$

It is important to note that for the particular case of a Pauli channel $\mathcal{P}$ the parameter r we have defined above is commonly called the "infidelity" of $\mathcal{P}$. Moreover, the term "error-rate" of $\mathcal{P}$ is sometimes defined to be the probability $r_{\mathcal{P}}$ that a non-identity Pauli operator is applied to the input state ρ. For the rest of the presentation we reserve the terms "error-rate" and "average error-rate" as we have defined above in Eq. (2.5). One can show using a Kraus representation of depolarizing channels (see Sec. A.2) that r and $r_{\mathcal{P}}$ are trivially related by dimensional factors via $r_{\mathcal{P}} = \frac{(d+1)r}{d}$. Indeed since

$$\frac{1}{d^2}\sum_{i=0}^{d^2-1} P_i \rho P_i = \frac{\mathbb{1}}{d} \tag{2.6}$$

we have,

$$\begin{aligned} \Lambda_d(\rho) &= p\rho + (1-p)\frac{\mathbb{1}}{d} \\ &= \left(p + \frac{1-p}{d^2}\right)\rho + (1-p)\left[\frac{1}{d^2}\sum_{i=1}^{d^2-1} P_i \rho P_i\right]. \end{aligned} \tag{2.7}$$

Therefore by definition of $r_{\mathcal{P}}$, and the fact that $r_{\mathcal{P}}$ is invariant under twirling,

$$\left(p + \frac{1-p}{d^2}\right)\rho + r_{\mathcal{P}}\left[\frac{1}{d^2-1}\sum_{i=1}^{d^2-1} P_i \rho P_i\right] = \left(p + \frac{1-p}{d^2}\right)\rho + (1-p)\left[\frac{1}{d^2}\sum_{i=1}^{d^2-1} P_i \rho P_i\right] \quad (2.8)$$

which implies $\frac{r_{\mathcal{P}}}{d^2-1} = \frac{1-p}{d^2}$. Hence,

$$r_{\mathcal{P}} = \frac{d^2 - 1(1-p)}{d^2} = \frac{(d+1)r}{d}. \quad (2.9)$$

The error-rate r is the figure of merit we want to be able to estimate experimentally. One can obtain estimates for p directly using the previously discussed methods of standard process tomography [33], ancilla-assisted/entanglement-assisted process tomography [7], ancilla-less selective tomography [126], symmetrization [47, 104] or Monte-Carlo methods [38, 50]. However as we noted before these methods suffer from various drawbacks which don't allow for a true estimate of p. For instance, the standard and ancilla-assisted tomography based schemes suffer from the unrealistic assumptions of negligible state-preparation and measurement errors and clean ancillary states/operations. Moreover, these methods require time resources exponential in n making them infeasible for even relatively small numbers of qubits. The ancilla-less selective tomography, symmetrization and Monte-Carlo methods also have the drawback of assuming negligible state-preparation and measurement error. However the advantages of these methods are that the average fidelity of each gate can be estimated and the schemes are efficient in n.

The experimentally relevant challenge therefore is to estimate p while relaxing the assumptions of clean state preparation, measurement and ancillary states/processes. Ideally, such a method should also scale efficiently with the number of qubits. As we now show, such an estimate can be obtained through our benchmarking protocol.

2.1.1 Protocol

In this section we give a detailed description of the randomized benchmarking protocol. In Sec. F.1, we provide a protocol which allows for the implementation of the benchmarking protocol in an experimental setting.

For a fixed sequence length $m \leq M-1$ (here M is the maximum sequence length), the benchmarking protocol consists of choosing K_m sequences of independent and identically

distributed uniformly random Clifford elements and calculating the average of the fidelity of the K_m sequences. One repeats this procedure for different values of m and fits the fidelity decay curve to the models we derive below. More precisely, the protocol is as follows,

Prepare an initial state $|\psi\rangle$ and perform the following steps:

Step 1. Fix $m \leq M-1$ and generate K_m sequences consisting of $m+1$ quantum operations (a discussion of how large K_m should be is given in Sec. 2.6). The first m operations are chosen uniformly at random from Clif_n and the $m+1$'th operation is uniquely determined as the inverse gate of the composition of the first m (see Fig. 2.1.1). By assumption, each operation $\mathcal{C}_{i_j}$ has some error, represented by $\Lambda_{i_j,j}$, and each sequence can be modelled by the operation,

$$\mathcal{S}_{\mathbf{i_m}} = \bigcirc_{j=1}^{m+1} \left(\Lambda_{i_j,j} \circ \mathcal{C}_{i_j}\right). \tag{2.10}$$

Here $\mathbf{i_m}$ is the m-tuple $(i_1, ..., i_m)$, i_{m+1} is uniquely determined by $\mathbf{i_m}$, and "$\circ$" represents composition.

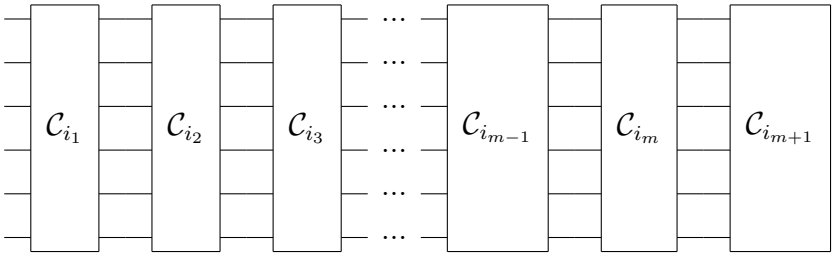

Figure 2.1: Experimental gate sequence: for each $j \in \{1, ..., m\}$, $\mathcal{C}_{i_j}$ is a randomly chosen Clifford and $\mathcal{C}_{i_{m+1}} = (\mathcal{C}_{i_m} \circ ... \circ \mathcal{C}_{i_1})^\dagger$ is the inverse gate.

Step 2. For each of the K_m sequences, measure the survival probability $\mathrm{Tr}[E_\psi \mathcal{S}_{\mathbf{i_m}}(\rho_\psi)]$. Here ρ_ψ is a (possibly mixed) quantum state that takes into account errors in preparing $|\psi\rangle\langle\psi|$ and E_ψ is the POVM element that takes into account measurement errors. In the ideal (noise-free) case

$$\rho_\psi = E_\psi = |\psi\rangle\langle\psi|. \tag{2.11}$$

Step 3. Average over the K_m random realizations to find the averaged sequence fidelity,

$$F_{\text{seq}}(m,\psi) = \text{Tr}[E_\psi \mathcal{S}_{K_m}(\rho_\psi)], \tag{2.12}$$

where

$$\mathcal{S}_{K_m} = \frac{1}{K_m}\sum_{\mathbf{i_m}} \mathcal{S}_{\mathbf{i_m}} \tag{2.13}$$

is the average sequence operation.

Step 4. Repeat Steps 1 through 3 for different values of m and fit the results for the averaged sequence fidelity (defined in Eq. (2.12)) to the model

$$\mathcal{F}_g^{(1)}(m,|\psi\rangle) = A_1 p^m + B_1 + C_1(m-1)(q-p^2)p^{m-2} \tag{2.14}$$

derived in Sec. 2.1.2. The coefficients A_1, B_1, and C_1 absorb the state preparation and measurement errors as well as the error on the final gate. The difference $q-p^2$ is a measure of the degree of gate-dependence in the errors, and p determines the average error-rate r according to the relation given by Eq. (2.5). In the case of gate-independent and time-independent errors the results will fit the simpler model

$$\mathcal{F}_g^{(0)}(m,|\psi\rangle) = A_0 p^m + B_0 \tag{2.15}$$

also derived in Sec. 2.1.2, where A_0 and B_0 absorb state preparation and measurement errors as well as the error on the final gate.

We note that for each m, in the limit of $K_m \to \infty$, $F_{\text{seq}}(m,\psi)$ converges to the exact (uniform) average, $\mathcal{F}_g(m,\psi)$, over all sequences,

$$\mathcal{F}_g(m,\psi) \;\; = \;\; \text{Tr}[E_\psi \mathcal{S}_m(\rho_\psi)] \tag{2.16}$$

where we define the exact average of the sequences to be,

$$\mathcal{S}_m = \frac{1}{|\mathrm{Clif}_n|^m} \sum_{(i_1,\ldots,i_m)} \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \ldots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}. \tag{2.17}$$

Hence the fitting functions by which we model the behavior of $F_{\mathrm{seq}}(m,\psi)$ are derived in terms of $\mathcal{F}_g(m,\psi)$ (this will be more clear in Sec. 2.1.2). Note that since $\mathcal{F}_g(m,\psi)$ is the uniform average over all sequences we can sum over each index in the above equation independently,

$$\mathcal{F}_g(m,\psi) = \frac{1}{|\mathrm{Clif}_n|^m} \sum_{i_1,\ldots,i_m} \mathrm{tr}\left(\Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m} \circ \mathcal{C}_{i_m} \circ \ldots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}(\rho_\psi) E_\psi\right). \tag{2.18}$$

Implicit in this equation is the assumption that the noise affecting each gate is independent of the noisy gates applied at earlier times in the sequence.

In order to prepare for the derivation of the above fitting models, we write $\mathcal{F}_g(m,\psi)$ in a more intuitive form. We first re-write $\Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m} \circ \mathcal{C}_{i_m} \circ \ldots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}$ by inductively defining new uniformly random gates from the Clifford group in the following manner:

1. Define $\mathcal{D}_{i_1} = \mathcal{C}_{i_1}$.

2. Define $\mathcal{D}_{i_2}$ uniquely by the equation $\mathcal{C}_{i_2} = \mathcal{D}_{i_2} \circ \mathcal{D}_{i_1}^\dagger$, ie.

$$\mathcal{D}_{i_2} = \mathcal{C}_{i_2} \circ \mathcal{C}_{i_1} = \bigcirc_{s=1}^{2} \mathcal{C}_{i_s}. \tag{2.19}$$

3. In general, for $j \in \{2,\ldots,m\}$, if $\mathcal{C}_{i_1}$,...,$\mathcal{C}_{i_j}$ and $\mathcal{D}_{i_1}$,...,$\mathcal{D}_{i_j}$ have been chosen, define $\mathcal{D}_{i_{j+1}}$ uniquely by the equation $\mathcal{C}_{i_{j+1}} = \mathcal{D}_{i_{j+1}} \circ \mathcal{D}_{i_j}{}^\dagger$, ie.

$$\mathcal{D}_{i_{j+1}} = \mathcal{C}_{i_{j+1}} \circ \ldots \circ \mathcal{C}_{i_1} = \bigcirc_{s=1}^{j+1} \mathcal{C}_{i_s}. \tag{2.20}$$

Note that if $j \neq k$, $\mathcal{C}_{i_j}$ and $\mathcal{C}_{i_k}$ are independent and so since the Clifford elements form a group, for each $j = 2,\ldots,m$, $\mathcal{D}_{i_j}$ is independent of $\mathcal{D}_{i_{j-1}}$. As well, summing over each i_j index runs over every Clifford element once and only once in $\mathcal{D}_{i_j}$.

We have created a new sequence $(\mathcal{D}_{i_1}, ..., \mathcal{D}_{i_m})$ from $(\mathcal{C}_{i_1}, ..., \mathcal{C}_{i_m})$ uniquely so that

$$\begin{aligned} \mathcal{S}_{\vec{i_m}} &= \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m} \circ \mathcal{C}_{i_m} \circ ... \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1} \\ &= \Lambda_{i_{m+1},m+1} \circ \mathcal{D}_{i_{m+1}} \circ \left({\mathcal{D}_{i_m}}^{\dagger} \circ \Lambda_{i_m,m} \circ \mathcal{D}_{i_m} \right) \circ ... \circ \left({\mathcal{D}_{i_1}}^{\dagger} \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1} \right). \end{aligned} \quad (2.21)$$

Since $\mathcal{C}_{i_{m+1}} = \mathcal{C}_{i_1}^{\dagger} \circ ... \circ \mathcal{C}_{i_m}^{\dagger}$ and $\mathcal{D}_{i_{m+1}} = \mathcal{C}_{i_{m+1}} \circ ... \circ \mathcal{C}_{i_1}$,

$$\mathcal{D}_{i_{m+1}} = \mathbb{1}. \quad (2.22)$$

Hence the $m+1$'th gate is effectively removed from the sequence in this change of variables and we have

$$\mathcal{S}_{\vec{i_m}} = \Lambda_{i_{m+1},m+1} \circ {\mathcal{D}_{i_m}}^{\dagger} \circ \Lambda_{i_m,m} \circ \mathcal{D}_{i_m} \circ ... \circ {\mathcal{D}_{i_1}}^{\dagger} \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1}. \quad (2.23)$$

2.1.2 Perturbative Expansion and the Fitting Models

We would like to develop fitting models for $\mathcal{F}_g(m, \psi)$ where the most general noise model allows for the noise to depend upon both the set of gates in Clif_n and time. We can estimate the behavior of $\mathcal{F}_g(m, \psi)$ by considering a perturbative expansion of each $\Lambda_{i,j}$ about the average Λ. We quantify the difference between $\Lambda_{i,j}$ and Λ by defining for all i, j,

$$\delta\Lambda_{i,j} = \Lambda_{i,j} - \Lambda. \quad (2.24)$$

Our approach will be valid provided the $\delta\Lambda_{i,j}$ are small perturbations from Λ in an average sense that is made precise in Sec. 2.2. More precisely, when the average variation of the perturbations is not too large, one can fit the experimental fidelity decay to a model that determines the average error per gate and provides a measure of the gate-dependence of the noise. Note that each $\delta\Lambda_{i,j}$ is a Hermiticity-preserving, trace-annihilating (see Sec. A.2) linear superoperator since it is the difference of trace-preserving, completely positive linear maps.

Using the change of variables $\mathcal{D}_{i_j} = \bigcirc_{s=1}^{j} \mathcal{C}_{i_s}$ described in Sec. 2.1.1 and expanding to first order we get,

$$
\begin{aligned}
\mathcal{S}_{\vec{i_m}} &= \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ ... \circ \Lambda_{i_j,j} \circ \mathcal{C}_{i_j} \circ ... \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1} \\
&= \Lambda_{i_{m+1},m+1} \circ {\mathcal{D}_{i_m}}^\dagger \circ \Lambda_{i_m,m} \circ \mathcal{D}_{i_m} \circ ... \circ {\mathcal{D}_{i_1}}^\dagger \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1} \\
&= \quad \Lambda \circ {\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \circ ... \circ {\mathcal{D}_{i_1}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} \\
&\quad + \delta\Lambda_{i_{m+1},m+1} \circ \left({\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ ... \circ \left({\mathcal{D}_{i_1}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + ... + \Lambda \circ \left({\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ ... \circ \left({\mathcal{D}_{i_j}}^\dagger \circ \delta\Lambda_{i_j,j} \circ \mathcal{D}_{i_j}\right) \circ ... \circ \left({\mathcal{D}_{i_1}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + ... + \Lambda \circ \left({\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ ... \circ \left({\mathcal{D}_{i_1}}^\dagger \circ \delta\Lambda_{i_1,1} \circ \mathcal{D}_{i_1}\right) \\
&\quad + O(\delta\Lambda^2_{i_j,j}).
\end{aligned} \tag{2.25}
$$

We define

$$
\mathcal{S}^{(0)}_{\vec{i_m}} := \Lambda \circ {\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \circ ... \circ {\mathcal{D}_{i_1}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}, \tag{2.26}
$$

$$
\begin{aligned}
\mathcal{S}^{(1)}_{\vec{i_m}} &:= \quad \delta\Lambda_{i_{m+1},m+1} \circ \left({\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ ... \circ \left({\mathcal{D}_{i_1}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + ... + \Lambda \circ \left({\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ ... \circ \left({\mathcal{D}_{i_j}}^\dagger \circ \delta\Lambda_{i_j,j} \circ \mathcal{D}_{i_j}\right) \circ ... \circ \left({\mathcal{D}_{i_1}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + ... + \Lambda \circ \left({\mathcal{D}_{i_m}}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ ... \circ \left({\mathcal{D}_{i_1}}^\dagger \circ \delta\Lambda_{i_1,1} \circ \mathcal{D}_{i_1}\right)
\end{aligned} \tag{2.27}
$$

and so on for higher order perturbation terms. Hence for each $j \in \{0, ..., m+1\}$, $\mathcal{S}^{(j)}_{\vec{i_m}}$ contains all j-body terms in the perturbative expansion.

Recalling the definition of $\mathcal{S}_m$ in Eq. (2.17), we also define for each order $k \in \{0, ..., m+1\}$,

$$
\mathcal{S}^{(k)}_m := \frac{1}{|\mathrm{Clif}_n|^m} \sum_{i_1,...,i_m} \mathcal{S}^{(k)}_{\vec{i_m}} \tag{2.28}
$$

and

$$\mathcal{F}_g^{(k)}(m,\psi) := \mathrm{tr}\left[\left(\sum_{j=0}^{k}\mathcal{S}_m^{(j)}\right)(\rho_\psi)E_\psi\right] \tag{2.29}$$

which represents the k'th order approximation to $\mathcal{F}_g(m,\psi)$. Hence,

$$\mathcal{S}_m = \sum_{k=0}^{m+1}\mathcal{S}_m^{(k)} \tag{2.30}$$

and

$$\begin{aligned}\mathcal{F}_g(m,\psi) &= \mathcal{F}_g^{(m+1)}(m,|\psi\rangle)\\ &= \mathrm{tr}\left[E_\psi\left(\sum_{j=0}^{m+1}\mathcal{S}_m^{(j)}\right)(\rho_\psi)\right]\\ &= \mathrm{tr}\left[E_\psi\mathcal{S}_m(\rho_\psi)\right].\end{aligned} \tag{2.31}$$

Zeroth Order Model

First, we look at the zeroth order fitting model $\mathcal{F}_g^{(0)}(m,|\psi\rangle)$ and note that $\mathcal{F}_g^{(0)}(m,|\psi\rangle)$ is exact in the case that the noise is independent of both the gate chosen and time (ie. $\Lambda_{i_j,j}=\Lambda$ for each time-step j and Clifford i_j). By independence of the $\mathcal{D}_{i_j}$ and the fact that averaging over the ensemble of realizations produces independent twirls which depolarize the m instances of Λ (see Sec. (C.2)) we get,

$$S_m^{(0)} = \Lambda\circ\Lambda_d\circ...\circ\Lambda_d = \Lambda\circ\left(\bigcirc_{j=1}^{m}\Lambda_d\right). \tag{2.32}$$

Since,

$$\begin{aligned}\Lambda \circ \left(\bigcirc_{j=1}^{m}\Lambda_d\right)(\rho_\psi) &= \Lambda\left(p^m\rho_\psi + (1-p^m)\frac{\mathbb{1}}{d}\right) \\ &= p^m\Lambda(\rho_\psi) + (1-p^m)\Lambda\left(\frac{\mathbb{1}}{d}\right) \end{aligned} \tag{2.33}$$

we get,

$$\begin{aligned}\mathcal{F}_g^{(0)}(m,|\psi\rangle) &= \mathrm{tr}\left(S_m^{(0)}(\rho_\psi)E_\psi\right) \\ &= \mathrm{tr}\left(\Lambda(\rho_\psi)E_\psi\right)p^m + \mathrm{tr}\left(\Lambda\left(\frac{\mathbb{1}}{d}\right)E_\psi\right)(1-p^m) \\ &= A_0p^m + B_0 \end{aligned} \tag{2.34}$$

where

$$A_0 := \mathrm{Tr}\left[\Lambda\left(\rho_\psi - \frac{\mathbb{1}}{d}\right)E_\psi\right] \tag{2.35}$$

and

$$B_0 := \mathrm{Tr}\left[\Lambda\left(\frac{\mathbb{1}}{d}\right)E_\psi\right]. \tag{2.36}$$

Hence, assuming the simplest (ideal) scenario where the noise operator is time and gate-independent, $\mathcal{F}_g(m,\psi) = \mathcal{F}_g^{(0)}(m,|\psi\rangle)$ decays exponentially in p.

First Order Model

To find $\mathcal{F}_g^{(1)}(m,|\psi\rangle)$ we note that in the definition of $S_{\vec{i}_m}^{(1)}$ given by Eq. (2.27) there are $\binom{m+1}{1} = m+1$ first-order perturbation terms which contain the gate dependence. First, we consider the $m-1$ terms with $j \in \{2,...,m\}$. For each such j, averaging over $i_1...i_m$ gives,

$$\frac{1}{|\mathrm{Clif}_n|^m}\sum_{i_1\ldots i_m}\Lambda\circ\left(\mathcal{D}_{i_m}{}^\dagger\circ\Lambda\circ\mathcal{D}_{i_m}\right)\circ\ldots\circ\left(\mathcal{D}_{i_j}{}^\dagger\circ\delta\Lambda_{i_j,j}\circ\mathcal{D}_{i_j}\right)\circ\ldots\circ\left(\mathcal{D}_{i_1}{}^\dagger\circ\Lambda\circ\mathcal{D}_{i_1}\right). \tag{2.37}$$

For these $m-1$ terms the main trick is to realize that we can re-expand $\mathcal{D}_{i_j}=\mathcal{C}_{i_j}\circ\mathcal{D}_{i_{j-1}}$ in order to depolarize the unitarily rotated perturbation $\mathcal{C}_{i_j}^\dagger\circ\Lambda_{i_j,j}\circ\mathcal{C}_{i_j}$ with the *twirling* operation $\frac{1}{|\mathrm{Clif}_n|}\sum_{i_{j-1}}\mathcal{D}_{i_{j-1}}^\dagger\cdot\mathcal{D}_{i_{j-1}}$. More precisely, the above can be written as,

$$\begin{aligned}\Lambda\circ\Lambda_d^{m-j} \quad &\circ \quad \left[\frac{1}{|\mathrm{Clif}_n|^2}\sum_{i_{j-1},i_j}\mathcal{D}_{i_{j-1}}{}^\dagger\circ\mathcal{C}_{i_j}^\dagger\circ\left(\Lambda_{i_j,j}-\Lambda\right)\circ\mathcal{C}_{i_j}\circ\Lambda\circ\mathcal{D}_{i_{j-1}}\right]\\ &\circ \quad \left[\frac{1}{|\mathrm{Clif}_n|^{j-2}}\sum_{i_{j-2},\ldots,i_1}\left(\mathcal{D}_{i_{j-2}}{}^\dagger\circ\Lambda\circ\mathcal{D}_{i_{j-2}}\right)\circ\ldots\circ\left(\mathcal{D}_{i_1}{}^\dagger\circ\Lambda\circ\mathcal{D}_{i_1}\right)\right]\\ &= \quad \Lambda\circ\Lambda_d^{m-j}\circ\left(\left(\mathcal{Q}_j\circ\Lambda\right)_d-\Lambda_d^2\right)\circ\Lambda_d^{j-2},\end{aligned} \tag{2.38}$$

where

$$\mathcal{Q}_j:=\frac{1}{|\mathrm{Clif}_n|}\sum_i\mathcal{C}_i^\dagger\circ\Lambda_{i,j}\circ\mathcal{C}_i \tag{2.39}$$

and the subscript "d" represents the depolarization of the operator within brackets. Using the fact that depolarizing channels commute we get that Eq.(2.38) can be written as,

$$\Lambda\circ\Lambda_d^{m-j}\circ\left(\left(\mathcal{Q}_j\circ\Lambda\right)_d-\Lambda_d^2\right)\circ\Lambda_d^{j-2}=\Lambda\circ\left(\left(\mathcal{Q}_j\circ\Lambda\right)_d-\Lambda_d^2\right)\circ\Lambda_d^{m-2}. \tag{2.40}$$

For the term with $j=1$, averaging over $i_1,\ldots,i_m$ gives a term of the form,

$$\Lambda\circ\Lambda_d^{m-1}\circ\frac{1}{|\mathrm{Clif}_n|}\sum_{i_1}\mathcal{D}_{i_1}{}^\dagger\circ\delta\Lambda_{i_1,1}\circ\mathcal{D}_{i_1}=\Lambda\circ\Lambda_d^{m-1}\circ(\mathcal{Q}_1-\Lambda_d), \tag{2.41}$$

where

$$\begin{aligned}\mathcal{Q}_1 &:= \frac{1}{|\mathrm{Clif}_n|}\sum_{i_1}\left(\mathcal{D}_{i_1}^\dagger \circ \Lambda_{i_1,1}\circ \mathcal{D}_{i_1}\right)\\ &= \frac{1}{|\mathrm{Clif}_n|}\sum_{i}\left(\mathcal{C}_{i}^\dagger \circ \Lambda_{i,1}\circ \mathcal{C}_{i}\right).\end{aligned} \tag{2.42}$$

Lastly for the term with $j = m+1$, averaging gives,

$$\begin{aligned}&\frac{1}{|\mathrm{Clif}_n|^m}\sum_{i_1\ldots i_m}\delta\Lambda_{i_{m+1},m+1}\circ\left({\mathcal{D}_{i_m}}^\dagger\circ\Lambda\circ\mathcal{D}_{i_m}\right)\circ\ldots\circ\left({\mathcal{D}_{i_1}}^\dagger\circ\Lambda\circ\mathcal{D}_{i_1}\right)\\ =\frac{1}{|\mathrm{Clif}_n|^{m-1}}&\sum_{i_1\ldots i_{m-1}}\left(\frac{1}{|\mathrm{Clif}_n|}\sum_{i_m}\delta\Lambda_{i_{m+1},m+1}\circ\left({\mathcal{D}_{i_m}}^\dagger\circ\Lambda\circ\mathcal{D}_{i_m}\right)\right)\circ\ldots\circ\left({\mathcal{D}_{i_1}}^\dagger\circ\Lambda\circ\mathcal{D}_{i_1}\right).\end{aligned} \tag{2.43}$$

Since Clif_n is a group, if $i_1,\ldots,i_{m-1}$ is fixed, averaging over the i_m index runs through every Clifford element with equal frequency in the $\mathcal{D}_{i_m}$ random variable. Since $\Lambda_{i_{m+1},m+1}$ is just the error associated with the gate $\mathcal{D}_{i_m}^\dagger$, $\frac{1}{|\mathrm{Clif}_n|}\sum_{i_m}\delta\Lambda_{i_{m+1},m+1}\circ\left({\mathcal{D}_{i_m}}^\dagger\circ\Lambda\circ\mathcal{D}_{i_m}\right)$ is independent of the $i_1,\ldots,i_{m-1}$ indices. Hence we can define

$$\begin{aligned}\mathcal{R}_{m+1} &:= \frac{1}{|\mathrm{Clif}_n|}\sum_{i_m}\Lambda_{i_{m+1},m+1}\circ\left({\mathcal{D}_{i_m}}^\dagger\circ\Lambda\circ\mathcal{D}_{i_m}\right)\\ &= \frac{1}{|\mathrm{Clif}_n|}\sum_{i}\Lambda_{i',m+1}\circ\left(\mathcal{C}_i^\dagger\circ\Lambda\circ\mathcal{C}_i\right)\end{aligned} \tag{2.44}$$

where $\Lambda_{i',m+1}$ denotes the error that arises when the Clifford operation $\mathcal{C}_i^\dagger$ is applied at final time-step $m+1$. Again, using the group property of Clif_n we have,

$$\mathcal{R}_{m+1} = \frac{1}{|\mathrm{Clif}_n|}\sum_{i}\Lambda_{i,m+1}\circ\left(\mathcal{C}_i\circ\Lambda\circ{\mathcal{C}_i}^\dagger\right). \tag{2.45}$$

This decoupling of $\mathcal{R}_{m+1}$ allows us to write,

$$\frac{1}{|\mathrm{Clif}_n|^{m-1}}\sum_{i_1\ldots i_{m-1}}\left(\frac{1}{|\mathrm{Clif}_n|}\sum_{i_m}\delta\Lambda_{i_{m+1},m+1}\circ\left(\mathcal{D}_{i_m}{}^{\dagger}\circ\Lambda\circ\mathcal{D}_{i_m}\right)\right)\circ\ldots\circ\left(\mathcal{D}_{i_1}{}^{\dagger}\circ\Lambda\circ\mathcal{D}_{i_1}\right)$$
$$=(\mathcal{R}_{m+1}-\Lambda\circ\Lambda_d)\circ\Lambda_d^{m-1}. \tag{2.46}$$

Hence combining Eq.'s (2.32),(2.40),(2.41) and (2.46) gives,

$$\begin{aligned}S_m^{(0)}+S_m^{(1)} &= \Lambda\circ\Lambda_d^m+(\mathcal{R}_{m+1}-\Lambda\circ\Lambda_d)\circ\Lambda_d^{m-1}\\ &\quad+\sum_{j=2}^{m}\Lambda\circ\left((\mathcal{Q}_j\circ\Lambda)_d-\Lambda_d^2\right)\circ\Lambda_d^{m-2}+\Lambda\circ\Lambda_d^{m-1}\circ(\mathcal{Q}_1-\Lambda_d) \qquad (2.47)\\ &= \mathcal{R}_{m+1}\circ\Lambda_d^{m-1}+\sum_{j=2}^{m}\left(\Lambda\circ(\mathcal{Q}_j\circ\Lambda)_d\circ\Lambda_d^{m-2}\right)+\Lambda\circ\Lambda_d^{m-1}\circ\mathcal{Q}_1-m\left(\Lambda\circ\Lambda_d^m\right).\end{aligned}$$

To calculate $\mathcal{F}_g^{(1)}(m,|\psi\rangle):=\mathrm{tr}\left[\left(S_m^{(0)}+S_m^{(1)}\right)(\rho_\psi)E_\psi\right]$ we have,

$$\mathrm{tr}\left(\mathcal{R}_{m+1}\circ\Lambda_d^{m-1}(\rho_\psi)E_\psi\right)=G_{1,m+1}p^{m-1}+H_{1,m+1}, \tag{2.48}$$

$$\mathrm{tr}\left(\Lambda\circ(\mathcal{Q}_j\circ\Lambda)_d\circ\Lambda_d^{m-2}(\rho_\psi)E_\psi\right)=A_0q_jp^{m-2}+B_0, \tag{2.49}$$

$$\mathrm{tr}\left(\Lambda\circ\Lambda_d^{m-1}\circ\mathcal{Q}_1(\rho_\psi)E_\psi\right)=A_{1,1}p^{m-1}+B_0, \tag{2.50}$$

$$\mathrm{tr}\left(\Lambda\circ\Lambda_d^m(\rho_\psi)E_\psi\right)=A_0p^m+B_0, \tag{2.51}$$

where

$$
\begin{aligned}
G_{1,m+1} &:= \mathrm{tr}\left(\mathcal{R}_{m+1}\left(\rho_\psi - \frac{\mathbb{1}}{d}\right)E_\psi\right), \\
H_{1,m+1} &:= \mathrm{tr}\left(\mathcal{R}_{m+1}\left(\frac{\mathbb{1}}{d}\right)E_\psi\right), \\
A_{1,1} &:= \mathrm{tr}\left(\Lambda\left(\mathcal{Q}_1\left(\rho_\psi\right) - \frac{\mathbb{1}}{d}\right)E_\psi\right),
\end{aligned} \tag{2.52}
$$

A_0 and B_0 are as given in Eq.s (2.35) and (2.36), and q_j is the depolarization parameter for $\left(\mathcal{Q}_j \circ \Lambda\right)_d$. Thus,

$$
\begin{aligned}
\mathcal{F}_g^{(1)}(m, |\psi\rangle) &= G_{1,m+1}p^{m-1} + H_{1,m+1} + \sum_{j=2}^{m}(A_0 q_j p^{m-2} + B_0) + A_{1,1}p^{m-1} + B_0 - m\left(A_0 p^m + B_0\right) \\
&= p^{m-1}\left(G_{1,m+1} + A_{1,1} - A_0 p\right) + (m-1)A_0 p^{m-2}\left(\frac{\sum_{j=2}^{m} q_j}{m-1} - p^2\right) + H_{1,m+1}.
\end{aligned} \tag{2.53}
$$

Finally, we can also re-write Eq. (2.53) as,

$$
\mathcal{F}_g^{(1)}(m, |\psi\rangle) = A_1(m)p^m + B_1(m) + C_1(m-1)(q(m) - p^2)p^{m-2} \tag{2.54}
$$

where,

$$
\begin{aligned}
A_1(m) &= \mathrm{Tr}\left[E_\psi \Lambda\left(\frac{\mathcal{Q}_1(\rho_\psi)}{p} - \rho_\psi + \frac{(p-1)\mathbb{1}}{pd}\right)\right] \\
&\quad + \mathrm{Tr}\left[E_\psi \mathcal{R}_{m+1}\left(\frac{\rho_\psi}{p} - \frac{\mathbb{1}}{pd}\right)\right] \\
B_1(m) &= \mathrm{Tr}\left[E_\psi \mathcal{R}_{m+1}\left(\frac{\mathbb{1}}{d}\right)\right] \\
C_1 &= \mathrm{Tr}\left[E_\psi \Lambda\left(\rho_\psi - \frac{\mathbb{1}}{d}\right)\right] \\
q(m) &= \frac{\sum_{j=2}^{m} q_j}{m-1},
\end{aligned} \tag{2.55}
$$

and q_j is the depolarizing parameter defined by

$$(\mathcal{Q}_j \circ \Lambda)_d(\rho) = q_j \rho + (1 - q_j)\frac{\mathbb{1}}{d}. \tag{2.56}$$

We write the first order model in the form of Eq. (2.54) because of its similarity to that of the zeroth order model given by Eq. (2.34). The difference between Eq.'s (2.54) and (2.34) is the $C_1(m-1)(q(m)-p^2)p^{m-2}$ term contained in Eq. (2.54), which can be thought of as a measure of the gate-dependence of the noise. Note that we have absorbed instances of p into the constant $A_1(m)$ under the assumption that $A_1(m)$ will remain independent of $B_1(m)$, C_1, and p under this redefinition. If this is not the case, then one should use the formula given by Eq. (2.53) for fitting.

Again, we see that the edge effects, state-preparation and measurement errors are embedded in the three coefficients $A_1(m)$, $B_1(m)$, and C_1. Note that the m dependence in $q(m)$ and the $A_1(m)$, and $B_1(m)$ coefficients due to the last gate disappears if the errors don't change as a function of time.

2.2 Neglecting Higher Orders

2.2.1 Bounding Higher Order Perturbation Terms

We would like to give conditions for when one is justified in terminating the perturbative expansion at some order k. The main idea, as expressed in Eq. (2.57) below, is to bound the size of the terms in $S_m^{(k+1)}$. The method we use for quantifying the size of a linear superoperator is the "$1 \to 1$" norm where the maximization is restricted to Hermitian inputs (see Sec. A.4.2). This norm is denoted by $\| \ \|_{1\to 1}^H$ and has the following useful properties:

- submultiplicativity for Hermiticity-preserving superoperators,
- unitary invariance,
- $\|\mathcal{E}\|_{1\to 1}^H \leq 1$ for any quantum operation $\mathcal{E}$.

Once we have presented the theory, we will discuss the motivation for using $\| \ \|_{1\to 1}^{H}$ as opposed to more familiar norms used in quantum information theory such as the diamond norm $\| \ \|_{\diamond}$.

From Sec. A.4.2 we have that,

$$\begin{aligned}
\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| &= \left|\mathrm{tr}\left[\left(\sum_{j=0}^{k+1} S_m^{(j)}\right)(\rho_\psi)E_\psi\right] - \mathrm{tr}\left[\left(\sum_{j=0}^{k} S_m^{(j)}\right)(\rho_\psi)E_\psi\right]\right| \\
&= \left|\mathrm{tr}\left[S_m^{(k+1)}(\rho_\psi)E_\psi\right]\right| \\
&\leq \|S_m^{(k+1)}\|_{1\to 1}^{H} \qquad (2.57)
\end{aligned}$$

and so bounding $S_m^{(k+1)}$ provides a bound for how much the k and $k+1$-order fidelities will differ. We first look at the case of stopping at first order, ie. $k = 1$. There are $\binom{m+1}{2} = \frac{(m+1)m}{2}$ second order perturbation terms in Eq. (2.25). Let us look at at a term with perturbations at j_1 and j_2 where without loss of generality we assume $j_2 > j_1$. Using the triangle inequality along with the properties listed above gives

$$\begin{aligned}
&\left\| \frac{1}{|\mathrm{Clif}_n|^m} \sum_{\vec{i_m}} \Lambda \circ ... \circ \mathcal{D}_{i_{j_2}}^\dagger \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}} \circ ... \circ \mathcal{D}_{i_{j_1}}^\dagger \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}} \circ ... \circ \mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} \right\|_{1\to 1}^{H} \\
&\qquad \leq \frac{1}{|\mathrm{Clif}_n|^m} \sum_{\vec{i_m}} \|\Lambda\|_{1\to 1}^{H} \left\|\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right\|_{1\to 1}^{H} ... \left\|\mathcal{D}_{i_{j_2}}^\dagger \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}}\right\|_{1\to 1}^{H} ... \\
&\qquad\qquad \left\|\mathcal{D}_{i_{j_1}}^\dagger \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}}\right\|_{1\to 1} ... \left\|\mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right\|_{1\to 1}^{H} \\
&= \left(\|\Lambda\|_{1\to 1}^{H}\right)^{m-1} \frac{1}{|\mathrm{Clif}_n|} \sum_{i_{j_2}} \left\|\mathcal{D}_{i_{j_2}}^\dagger \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}}\right\|_{1\to 1}^{H} \frac{1}{|\mathrm{Clif}_n|} \sum_{i_{j_1}} \left\|\mathcal{D}_{i_{j_1}}^\dagger \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}}\right\|_{1\to 1}^{H} \\
&\qquad \leq \frac{1}{|\mathrm{Clif}_n|} \sum_{i_{j_2}} \left\|\mathcal{D}_{i_{j_2}}^\dagger \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}}\right\|_{1\to 1}^{H} \frac{1}{|\mathrm{Clif}_n|} \sum_{i_{j_1}} \left\|\mathcal{D}_{i_{j_1}}^\dagger \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}}\right\|_{1\to 1}^{H} \\
&\qquad\qquad = \gamma_{j_2}\gamma_{j_1} \qquad (2.58)
\end{aligned}$$

The first inequality in Eq. (2.2.1) follows from the triangle inequality and submultiplicativity, the proceeding equality follows from unitary invariance, the second inequality follows

from the third property of $\|\ \|^H_{1\to 1}$ listed above, and the second (last) equality follows from defining the time-dependent variation in the noise,

$$\gamma_j := \frac{1}{|\mathrm{Clif}_n|}\sum_i \|\Lambda_{i,j} - \Lambda\|^H_{1\to 1}\,. \tag{2.59}$$

Summing over all j_1, j_2 with $j_2 > j_1$ gives,

$$\begin{aligned}
\left\|S^{(2)}_m\right\|^H_{1\to 1} &= \left\| \frac{1}{|\mathrm{Clif}_n|^m}\sum_{\vec{i_m}} S^{(2)}_{\vec{i_m}} \right\|^H_{1\to 1} \\
&= \Big\| \frac{1}{|\mathrm{Clif}_n|^m}\sum_{\vec{i_m}}\sum_{j_2>j_1} \Lambda\circ\mathcal{D}^\dagger_{i_m}\circ\Lambda\circ\mathcal{D}_{i_m}\circ ...\circ\mathcal{D}^\dagger_{i_{j_2}}\circ\delta\Lambda_{i_{j_2}}\circ\mathcal{D}_{i_{j_2}}\circ ... \\
&\qquad\qquad \circ\mathcal{D}^\dagger_{i_{j_1}}\circ\delta\Lambda_{i_{j_1}}\circ\mathcal{D}_{i_{j_1}}\circ ...\circ\mathcal{D}^\dagger_{i_1}\circ\Lambda\circ\mathcal{D}_{i_1}\Big\|^H_{1\to 1} \\
&\le \sum_{j_2>j_1}\Big\| \frac{1}{|\mathrm{Clif}_n|^m}\sum_{\vec{i_m}} \Lambda\circ\mathcal{D}^\dagger_{i_m}\circ\Lambda\circ\mathcal{D}_{i_m}\circ ...\circ\mathcal{D}^\dagger_{i_{j_2}}\circ\delta\Lambda_{i_{j_2}}\circ\mathcal{D}_{i_{j_2}}\circ ... \\
&\qquad\qquad \circ\mathcal{D}^\dagger_{i_{j_1}}\circ\delta\Lambda_{i_{j_1}}\circ\mathcal{D}_{i_{j_1}}\circ ...\circ\mathcal{D}^\dagger_{i_1}\circ\Lambda\circ\mathcal{D}_{i_1}\Big\|^H_{1\to 1} \\
&\le \sum_{j_2>j_1}\gamma_{j_2}\gamma_{j_1}.
\end{aligned} \tag{2.60}$$

In terms of the fidelity we thus have from Eq.'s (2.57) and (2.60),

$$\left|\mathcal{F}^{(2)}_g(m,|\psi\rangle) - \mathcal{F}^{(1)}_g(m,|\psi\rangle)\right| \le \sum_{j_2>j_1}\gamma_{j_2}\gamma_{j_1}. \tag{2.61}$$

Hence when

$$\sum_{j_2>j_1}\gamma_{j_2}\gamma_{j_1} \ll 1 \tag{2.62}$$

we can stop at first order. Note that if the noise is time-independent then so is γ and we have,

$$\sum_{j_2>j_1} \gamma^2 = \frac{(m+1)m}{2}\gamma^2 \tag{2.63}$$

which gives,

$$\left|\mathcal{F}_g^{(2)}(m,|\psi\rangle) - \mathcal{F}_g^{(1)}(m,|\psi\rangle)\right| \leq \frac{(m+1)m}{2}\gamma^2. \tag{2.64}$$

Hence one is justified in stopping at first order in this case when

$$\gamma^2 << \frac{2}{(m+1)m}. \tag{2.65}$$

It is straightforward to show that bounds on higher order terms go as

$$\left\|S_m^{(k)}\right\|_{1\to 1}^H \leq \sum_{j_k>\ldots>j_1} \gamma_{j_k}\ldots\gamma_{j_1} \tag{2.66}$$

so that the difference between the k and $k+1$-order fidelities is bounded by,

$$\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| \leq \sum_{j_k>\ldots>j_1} \gamma_{j_k}\ldots\gamma_{j_1}. \tag{2.67}$$

Hence if

$$\sum_{j_k>\ldots>j_1} \gamma_{j_k}\ldots\gamma_{j_1} << 1 \tag{2.68}$$

we can stop at k'th order. Again if the noise is time-independent,

$$\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| \leq \binom{m+1}{k}\gamma^k \tag{2.69}$$

which implies one can stop at k'th order in this case if

$$\binom{m+1}{k}\gamma^k << 1. \tag{2.70}$$

Note that since

$$\binom{m+1}{k} \leq \frac{(m+1)^k}{k!} \tag{2.71}$$

we have,

$$\binom{m+1}{k}\gamma^k \leq \frac{((m+1)\gamma)^k}{k!}. \tag{2.72}$$

Therefore an easier condition to check whether one is justified in stopping at k'th order is if

$$\gamma^k << \frac{k!}{(m+1)^k}. \tag{2.73}$$

We now discuss our motivation for using $\| \ \|_{1\to 1}^H$ as opposed to more familiar norms for distinguishing superoperators, such as the diamond norm. For any superoperator norm $\| \ \|$ that satisfies the properties listed above, the following inequality holds,

$$|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)| \leq \binom{m+1}{k}\gamma^k \tag{2.74}$$

where,

$$\gamma := \frac{1}{|\mathrm{Clif}_n|}\sum_i \|\Lambda_i - \Lambda\| \tag{2.75}$$

and for simplicity we have assumed time-independent noise.

This implies that in order to give the tightest bound on the fidelity difference in Eq. (2.74), we would like to find the norm $\| \cdot \|$ that provides the smallest value of γ.

The diamond norm $\|\cdot\|_\diamond$ is a candidate however by Eq. (A.80) $\|\ \|_{1\to 1}^H$ is much weaker than $\|\cdot\|_\diamond$. Therefore γ associated with $\|\ \|_{1\to 1}^H$ will be much smaller than γ associated with $\|\cdot\|_\diamond$, providing a tighter bound in Eq. (2.74).

2.3 Case Where Benchmarking Fails

There is a highly unrealistic, yet simple to describe, case where benchmarking fails. Suppose the noise is time-independent and for each i, $\Lambda_i = \mathcal{C}_i^\dagger$. Then $F_g(m,\psi) = 1$ for every m even though there is substantial error on each $\mathcal{C}_i$. The key point to note here is that the noise is highly dependent on the gate chosen and so we expect that the sufficient condition derived above for ignoring higher order terms will not be satisfied (ie. γ in this example will be far from 0). To see that this is the case, note that since Clif_n is a unitary 2-design it is also a unitary 1-design. Hence since Clif_n is †-closed,

$$\begin{aligned}\frac{1}{|\text{Clif}_n|}\sum_{i=1}^{|\text{Clif}_n|}\Lambda_i &= \frac{1}{|\text{Clif}_n|}\sum_{i=1}^{|\text{Clif}_n|}\mathcal{C}_i^\dagger \\ &= \frac{1}{|\text{Clif}_n|}\sum_{i=1}^{|\text{Clif}_n|}\mathcal{C}_i \\ &= \Omega \end{aligned} \tag{2.76}$$

where Ω is the totally depolarizing channel mapping every input state to the maximally mixed state $\frac{\mathbb{1}}{d}$. Therefore,

$$\|\Lambda_i - \Lambda\|_{1\to 1}^H = \|\mathcal{C}_i^\dagger - \Omega\|_{1\to 1}^H. \tag{2.77}$$

Now the value of $\|\Lambda_i - \Lambda\|_{1\to 1}^H$ is achieved at a pure state [142] and for any pure state $|\psi\rangle$,

$$(\Lambda_i - \Lambda)(|\psi\rangle\langle\psi|) = C_i^\dagger|\psi\rangle\langle\psi|C_i - \frac{\mathbb{1}}{d}. \tag{2.78}$$

Hence if $|\phi\rangle$ is a pure state at which the value of $\|\Lambda_i - \Lambda\|_{1\to 1}^H$ is achieved,

$$\begin{aligned} \|\Lambda_i - \Lambda\|_{1\to 1}^H &= \left\| C_i^\dagger |\phi\rangle\langle\phi| C_i - \frac{\mathbb{1}}{d} \right\|_1 \\ &= 1 - \frac{1}{d} + (d-1)\frac{1}{d} \\ &= \frac{2(d-1)}{d}. \end{aligned} \tag{2.79}$$

Therefore in this case,

$$\begin{aligned} \gamma &= \frac{1}{|\mathrm{Clif}_n|} \sum_i \|\Lambda_i - \Lambda\|_{1\to 1}^H \\ &= \frac{2(d-1)}{d} \geq 1 \end{aligned} \tag{2.80}$$

and so our sufficient condition is not satisfied as expected.

It is important to note that one can devise tests for when such a pathological case occurs. One simple test is given as follows: If the input is a stabilizer state $|\psi\rangle$ then choose Clifford elements $\mathcal{C}_i$ that map $|\psi\rangle$ to an orthogonal state in a measurement basis containing $|\psi\rangle$. For each i, apply $\mathcal{C}_i$ to $|\psi\rangle$ and perform the measurement. For small noise strength the output of the measurement should almost never be $|\psi\rangle$, however if the noise is something close to the inverse of the gate the measurement result will be $|\psi\rangle$ with high probability.

2.4 State Preparation and Measurement Errors

In this section we analyze the effect of state preparation and measurement errors on the benchmarking protocol. The main result is that these errors can be ignored in almost any situation of practical relevance. The key point is that one can obtain an estimate for the depolarizing parameter p as long as the fidelity decay curve is not constant. Thus, since state-preparation and measurement errors are accounted for in A_0 and B_0, the protocol is robust against any state preparation or measurement errors unless these errors conspire to create a constant fidelity curve.

For simplicity of the discussion let us assume the gate-dependence of the noise is weak enough so that the zeroth order expression given in Eq. (2.34) is a valid model for the

fidelity decay curve. It is straightforward to characterize exactly when the fidelity curve is constant. Indeed, from Eq. (2.34) an exponential decay occurs if and only if A_0 is non-zero and p lies in $(0, 1)$. Hence no decay occurs if and only if one of $p = 0$, $p = 1$ or $A_0 = 0$ occurs. We look at each case separately.

$\underline{p = 0}$: This occurs if and only if Λ is the totally depolarizing channel and in this case the fidelity is constant at $B_0 = \frac{\mathrm{tr}(E_\psi)}{d} \leq \frac{1}{d}$. Since we have assumed small gate-dependence of the noise, this case is only possible if most of the errors are approximately centred around the totally depolarizing channel with little variation. This situation is of little practical relevance since the gate operations being characterized are usually reasonably precise.

$\underline{p = 1}$: This case corresponds to Λ being the identity channel which means all gates are perfect. Again, in practice this situation is unlikely as the implementation of any gate will have some associated error. Note that in this case the fidelity is equal to $A_0 + B_0$ which is just $\mathrm{tr}\left[\Lambda(\rho_\psi)E_\psi\right] = \mathrm{tr}(\rho_\psi E_\psi)$. Hence the constant decay curve is a measure of the overlap between the imperfect input state and imperfect POVM element.

$\underline{A_0 = 0}$: The case $A_0 = 0$ occurs if and only if

$$\mathrm{tr}(E_\psi \Lambda(\rho_\psi)) = \mathrm{tr}\left(E_\psi \Lambda\left(\frac{\mathbb{1}}{d}\right)\right). \tag{2.81}$$

Thus $\Lambda(\rho_\psi)$ and $\Lambda\left(\frac{\mathbb{1}}{d}\right)$ have the same probability of producing the output "ψ" from the measurement. Since gates are reasonably precise in practice, this situation occurs when at least one of state preparation or measurement has substantial error. Note that the fidelity will be equal to B_0 in this case and so can take any value in $[0, 1]$.

From the above three cases, the only one that depends upon state preparation or measurement errors is the case $A_0 = 0$. Since this case occurs when at least one of state preparation or measurement has substantial error it is unlikely to arise in practice. This discussion shows that a constant fidelity decay curve can only occur in extreme cases and so it is usually safe to assume the protocol is independent of state preparation and measurement errors.

2.5 Average Error Rate and the Diamond Norm

It is useful to draw connections between the average error rate r between Λ and $\mathcal{I}$ and more relevant measures of error used in fault-tolerance, such as the diamond norm between

Λ and $\mathcal{I}$ (see Sec. A.4.2 for the definition of the diamond norm $\|\cdot\|_\diamond$). In general an exact relationship will be impossible to obtain, however we show that in certain cases that are relevant in many scenarios, one being fault-tolerance, we can obtain such a relationship. First we give a new proof of a previously established result [125] for calculating the diamond norm distance between generalized Pauli channels. The proof presented here illustrates how one can apply the methods of semidefinite programming to calculate the diamond norm distance between quantum channels [144]. Ideally, this proof technique could be used to either explicitly calculate or place bounds on the diamond norm distance between more general classes of quantum channels. This could allow for obtaining relationships between r and the diamond norm distance which hold in more general cases.

2.5.1 Calculating the Diamond Norm Distance Between Generalized Pauli Channels

The key to obtaining useful relationships between the error rate and diamond norm is the following theorem:

Theorem 1. *Suppose $\mathcal{E}_1$ and $\mathcal{E}_2$ are Pauli channels, or more generally any channels with Kraus operators given by an orthogonal basis of unitary operators $\{P_i\}_{i=0}^{d^2}$ satisfying $\mathrm{tr}\,(P_i P_j) = d\delta_{i,j}$ (which we call generalized Pauli channels),*

$$\mathcal{E}_1(\rho) = \sum_{i=0}^{d^2-1} q_i P_i \rho P_i^\dagger \tag{2.82}$$

$$\mathcal{E}_2(\rho) = \sum_{i=0}^{d^2-1} r_i P_i \rho P_i^\dagger. \tag{2.83}$$

Define the vector $\vec{v}$ of length d^2 by

$$v_i = q_i - r_i \tag{2.84}$$

for all $i \in \{0, ..., d^2 - 1\}$. Then,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \|\vec{v}\|_1 = \sum_{i=0}^{d^2-1} |v_i|. \tag{2.85}$$

Proof. To prove Eq. (2.85) using the semidefinite program in [144] first note that $\Phi = \mathcal{E}_1 - \mathcal{E}_2$ has action,

$$\Phi(\rho) = \sum_{i=0}^{d^2-1} (q_i - r_i) P_i \rho P_i^\dagger. \tag{2.86}$$

The semidefinite program has the following primal and dual problems:

Primal problem: Maximize $\langle C(\Phi), W \rangle$ subject to

- $W \leq \mathbb{1}_d \otimes \rho$,
- $W \in \mathrm{Pos}\left(L\left(\mathbb{C}^d \otimes \mathbb{C}^d\right)\right)$,
- $\rho \in \mathcal{D}\left(L\left(\mathbb{C}^d\right)\right)$.

Dual problem: Minimize $\|\mathrm{tr}_1(Z)\|_\infty$ subject to

- $Z \geq C(\Phi)$,
- $Z \in \mathrm{Pos}(L\left(\mathbb{C}^d \otimes \mathbb{C}^d\right))$,

where $C(\Phi)$ is the Choi matrix [29] of Φ (see Sec. A.2). If α and β are the solutions to the primal and dual problems then the case that $\alpha = \beta$ is called *strong duality*. It is shown in [144] that the above semidefinite program always has the property of strong duality and the solution to the program is $\alpha = \frac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond$. Note that in general, it is the case that $\alpha \leq \beta$.

By definition,

$$
\begin{aligned}
C(\Phi) &= d\Phi \otimes \mathcal{I}(|\psi_0\rangle\langle\psi_0|) \\
&= d \sum_{i=0}^{d^2-1} (q_i - r_i) P_i \otimes \mathbb{1} |\psi_0\rangle\langle\psi_0| P_i^\dagger \otimes \mathbb{1}
\end{aligned} \tag{2.87}
$$

where $|\psi_0\rangle$ is as defined in Sec. A.1.1. Note that the set of states $\{|\psi_i\rangle\}_{i=0}^{d^2-1}$ defined by

$$
|\psi_i\rangle := (P_i \otimes \mathbb{1}) |\psi_0\rangle \tag{2.88}
$$

forms an orthonormal basis that consists of maximally entangled states in $\mathbb{C}^d \otimes \mathbb{C}^d$. We call this basis the generalized Bell basis (GBB) and $C(\Phi)$ is diagonal when written in GBB with diagonal elements (eigenvalues) $d(q_i - r_i)$. Let Π_+ denote the projector onto the eigenspace of $C(\Phi)$ with non-negative eigenvalues and Π_- denote the projector onto the eigenspace with negative eigenvalues.

For the primal problem let $W = \frac{\Pi_+}{d}$ and $\rho = \frac{\mathbb{1}}{d}$. Then

$$
\begin{aligned}
\langle C(\Phi), W \rangle &= \sum_{k: q_k - r_k \geq 0} q_k - r_k \\
&= \frac{1}{2} \sum_k |q_k - r_k| \\
&= \frac{1}{2} \|\vec{v}\|_1.
\end{aligned} \tag{2.89}
$$

Thus,

$$
\alpha \geq \frac{1}{2} \|\vec{v}\|_1. \tag{2.90}
$$

For the dual problem take $Z = d\Pi_+ C(\Phi) \Pi_+$ and note that

$$
Z = \sum_{k: q_k - r_k \geq 0} (q_k - r_k) |\psi_k\rangle\langle\psi_k|. \tag{2.91}
$$

Hence, $Z \geq C(\Phi)$. Moreover,

$$\mathrm{tr}_1(Z) = d\left(\sum_{k:q_k-r_k\geq 0} q_k - r_k\right)\frac{\mathbb{1}}{d} \tag{2.92}$$

and so

$$\begin{aligned}\|\mathrm{tr}_1(Z)\|_\infty &= \left(\sum_{k:q_k-r_k\geq 0} q_k - r_k\right) \\ &= \frac{1}{2}\|\vec{v}\|_1. \end{aligned} \tag{2.93}$$

Thus $\alpha \leq \frac{1}{2}\|\vec{v}\|_1$ which implies $\alpha = \frac{1}{2}\|\vec{v}\|_1$ and $\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \|\vec{v}\|_1$ as desired.

□

There is a simple corollary to Eq. (2.85) in the case of depolarizing channels.

Corollary 1. *Suppose $\mathcal{E}_1$ and $\mathcal{E}_2$ are depolarizing channels of the form*

$$\mathcal{E}_1(\rho) = p_1\rho + (1-p_1)\frac{\mathbb{1}}{d} \tag{2.94}$$

and

$$\mathcal{E}_2(\rho) = p_2\rho + (1-p_2)\frac{\mathbb{1}}{d}. \tag{2.95}$$

Then,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \frac{2|p_1-p_2|(d^2-1)}{d^2}. \tag{2.96}$$

Proof. To see this note that in this case,

$$\begin{aligned} q_0 &= \frac{(d+1)\overline{F_{\mathcal{E}_1,\mathcal{I}}} - 1}{d} \\ &= \frac{(d+1)\left(p_1 + \frac{1-p_1}{d}\right) - 1}{d} \\ &= \frac{(d^2-1)p_1 + 1}{d^2} \end{aligned} \tag{2.97}$$

and similarly,

$$r_0 = \frac{(d^2-1)p_2 + 1}{d^2}. \tag{2.98}$$

Thus for every $1 \leq i \leq d^2 - 1$,

$$q_i = \frac{1-q_0}{d^2-1} = \frac{1-p_1}{d^2} \tag{2.99}$$

and

$$r_i = \frac{1-r_0}{d^2-1} = \frac{1-p_2}{d^2}. \tag{2.100}$$

So,

$$\begin{aligned} \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond &= \|v\|_1 \\ &= |q_0 - r_0| + \sum_{i=1}^{d^2-1} |q_i - r_i| \\ &= \left|\frac{(d^2-1)p_1 + 1}{d^2} - \left(\frac{(d^2-1)p_2 + 1}{d^2}\right)\right| + (d^2-1)\left|\frac{1-p_1}{d^2} - \left(\frac{1-p_2}{d^2}\right)\right| \\ &= \frac{2(d^2-1)|p_1 - p_2|}{d^2}. \end{aligned} \tag{2.101}$$

□

2.5.2 Relating the Diamond Norm With the Error Rate Obtained From Benchmarking

Now suppose that $\mathcal{E}_2 = \mathcal{I}$ in Eq. (2.85). Then, $r_0 = 1$ and for every $1 \leq i \leq d^2 - 1$, $r_i = 0$. Hence in this case,

$$\begin{aligned} \|\mathcal{E}_1 - \mathcal{I}\|_\diamond &= \|\vec{v}\|_1 \\ &= |q_0 - 1| + 1 - q_0 \\ &= 2(1 - q_0). \end{aligned} \tag{2.102}$$

We know that q_0 is related to the average fidelity of $\mathcal{E}_1$, $\overline{F_{\mathcal{E}_1,\mathcal{I}}}$, by

$$\overline{F_{\mathcal{E}_1,\mathcal{I}}} = \frac{q_0 d + 1}{d + 1} \tag{2.103}$$

and so,

$$\|\mathcal{E}_1 - \mathcal{I}\|_\diamond = \frac{2(d+1)(1 - \overline{F_{\mathcal{E}_1,\mathcal{I}}})}{d}. \tag{2.104}$$

Therefore in the case of randomized benchmarking (where we define the error rate $r = 1 - \overline{F_{\Lambda,\mathcal{I}}}$) if Λ is a generalized Pauli channel then r and $\|\Lambda - \mathcal{I}\|_\diamond$ are related by,

$$\|\Lambda - \mathcal{I}\|_\diamond = \frac{2(d+1)r}{d}. \tag{2.105}$$

Eq. (2.105) implies that when Λ is a Pauli channel we can, in principle, deduce the exact value of $\|\Lambda - \mathcal{I}\|_\diamond$ via the scalable and robust randomized benchmarking protocol presented in this chapter. As mentioned previously, the diamond norm is usually the standard figure of merit used to characterize threshold values in fault-tolerant computation [78, 2], and is

a much stronger measure of the distance between quantum operations than the average fidelity. The best known relationship between the diamond norm and average fidelity for arbitrary Λ is given in [10] where their results imply that in general,

$$\|\Lambda - \mathcal{I}\|_\diamond \leq 4\sqrt{d(d+1)r}. \tag{2.106}$$

Clearly the above bound does not scale well in d and quickly becomes a poor measure of the size of $\|\Lambda - \mathcal{I}\|_\diamond$. In this sense it may seem somewhat surprising that an exact relationship between r and $\|\Lambda - \mathcal{I}\|_\diamond$ can be obtained for the case of Λ being a Pauli channel.

A key point to observe is that since Λ is the average of $2^{O(n^2)}$ noise operators, one expects that for a large class of error models Λ_i, Λ has no strong polarization preference, ie. it is close to a depolarizing channel. As depolarizing channels are Pauli channels, this would imply that estimating r gives a very good approximation of $\|\Lambda - \mathcal{I}\|_\diamond$. Of course there are obvious exceptions to this case, for instance when for each i, $\Lambda_i = \Lambda$ where Λ is say a small unitary rotation.

2.6 Scalability of the Protocol

In this section we provide a proof of the scalability of the RB protocol. First, we note that the size of the Clifford group scales as $2^{O\left(n^2\right)}$ and so the number of sequences of length m scales as $2^{mO\left(n^2\right)}$. There are four main points to discuss for scalability:

1. Sequence length: Since the number of sequences of length m scales as $2^{mO\left(n^2\right)}$, averaging over *all* sequences for each m is clearly inefficient.

2. Uniform sampling: Since the size of the Clifford group scales as $2^{O\left(n^2\right)}$, sampling directly from a list of all Clifford elements becomes impossible for large n (just writing down every element is inefficient in n).

3. Decomposing Clifford operations: In practice, one can only implement a generating set for the Clifford group. Hence even if random sampling can be accomplished there must be a scalable method for decomposing each Clifford into a sequence of generators.

4. Deterministic final gate: The $m+1$'th Clifford operation is deterministically chosen from the first m random Clifford elements. One needs to verify that determining this final gate is a scalable process.

We now describe how to overcome each of these potential obstacles. The theory on the symplectic representation of the Clifford group given in Sec. E.1 is important, but not necessary, for the discussion. The reader unfamiliar with the symplectic representation is encouraged to read through Sec. E.1 for an introduction with examples.

Solution to 1: From Eq. (2.18), $\mathcal{F}_g(m,\psi)$ is the uniform average of the random variable

$$\begin{aligned}\mathcal{F}_g^{\vec{i_m}}(m,|\psi\rangle) &:= \mathrm{tr}\left(S_{\vec{i_m}}(\rho_\psi)E_\psi\right) \\ &= \mathrm{tr}\Big(\Lambda_{i_{m+1},m+1}\circ\mathcal{C}_{i_{m+1}}\circ...\circ\Lambda_{i_1,1}\circ\mathcal{C}_{i_1}(\rho_\psi)E_\psi\Big) \end{aligned} \tag{2.107}$$

over $|\mathrm{Clif}_n|^m$ sequences $(i_1,...,i_m)$. The benchmarking protocol requires choosing a sequence at random, evaluating the above fidelity, repeating for many sequences, and taking the average of the results.

Let $S_K(m,|\psi\rangle)$be the normalized K-fold sum of the random variable $\mathcal{F}_g^{\vec{i_m}}(m,|\psi\rangle)$ and note that $\mathbb{E}[S_K(m,|\psi\rangle)] = \mathcal{F}_g(m,\psi)$ where "$\mathbb{E}$" represents "expectation value". A probablistic bound on $|S_K(m,|\psi\rangle) - \mathcal{F}_g(m,\psi)|$ is given by Höeffding's inequality,

$$\begin{aligned}\mathbb{P}\left(|S_K(m,|\psi\rangle) - \mathcal{F}_g(m,\psi)| \geq \epsilon\right) &\leq 2e^{\frac{-2(K\epsilon)^2}{K(b-a)^2}} \\ &= 2e^{\frac{-2K\epsilon^2}{(b-a)^2}} \end{aligned} \tag{2.108}$$

where ϵ represents the accuracy of the estimate, $[a,b]$ is the range of $\mathcal{F}_g^{\vec{i_m}}(m,|\psi\rangle)$, and "$\mathbb{P}$" represents "probability". Since $\mathcal{F}_g^{\vec{i_m}}(m,|\psi\rangle)$ is a fidelity it must lie in $[0,1]$ (in reality it will lie in a much smaller interval, for now we continue to assume it lies in some $[a,b]\subseteq[0,1]$). Suppose we want

$$\mathbb{P}\left(|S_K(m,|\psi\rangle) - \mathcal{F}_g(m,\psi)| \geq \epsilon\right) \leq \delta \tag{2.109}$$

where $1-\delta$ represents the desired confidence level. We can find how many trials one needs to perform to obtain accuracy ϵ with confidence $1-\delta$ by setting $\delta = 2e^{\frac{-2K\epsilon^2}{(b-a)^2}}$ and solving for K,

$$K = \frac{\ln\left(\frac{2}{\delta}\right)(b-a)^2}{2\epsilon^2}. \tag{2.110}$$

Note that K is explicitly independent of m and n which provides a solution to Problem 1.

It is instructive to obtain an estimate of the size of K for realistic parameter values of δ and ϵ. Since $1-\delta$ represents our desired confidence level we set $\delta = 0.05$. Fault-tolerance provides a wide range for the error tolerance of a physical (0-level) gate in the fault-tolerant construction. The value of the error tolerance depends on both the coding scheme as well as the noise model and typical values lie somewhere between 10^{-6} and 10^{-2}. Let us assume that the physical gates have errors on the order of 10^{-4}. Intuitively, since the fidelity curve decays in sequence length it is reasonable to assume that ϵ can be relaxed as m grows large. Similarly, $b-a$ can be assumed to be relatively small for small values of m but will converge to $1-\frac{1}{d}$ as m grows large. As a result both $b-a$ and ϵ have an *implicit* dependence on m and this implicit dependence is advantageous when choosing ϵ for large values of m. Let us assume $m = 100$ and a fidelity decay curve that is well-approximated by an exponential. Then we expect fidelity values on the order of 0.99 at this value of m and so we take $\epsilon = 10^{-3}$, $b-a = 0.2$. With these values for ϵ, δ and $b-a$ we get the number of trials required is,

$$\begin{aligned} K &= \frac{\ln\left(\frac{2}{0.05}\right)(0.2)^2}{2(10^{-3})^2} \\ &\sim 7 \times 10^4. \end{aligned} \tag{2.111}$$

While this number is large it is independent of n and thus compares favourably with quantum process tomography which scales as 16^n. As a direct comparison, performing process tomography on a 4 qubit system already requires 65536 measurements.

Solution to 2:

For the second problem we present a scalable method for uniform sampling from the full Clifford group which utilizes the symplectic representation of the Clifford group (see Ref's [39, 41]). Sec E.1 contains an introduction to the symplectic representation with various examples. Since the Clifford group is the normalizer of the Pauli group, every Clifford element $\mathcal{Q}$ is completely determined by its action under conjugation on the Pauli group. In particular, since the Pauli group is generated by the set of all X_i and Z_i (the

label i refers to X or Z being in the i'th position with identity operators elsewhere), $\mathcal{Q}$ is completely determined by its action on this set. In the symplectic representation this corresponds to each $\mathcal{Q}$ being associated uniquely to a $2n$ by $2n$ binary symplectic matrix C and length $2n$ binary vector h which records negative signs in the images of X_i and Z_i. The only constraints on $\mathcal{Q}$ are that commutation relations and Hermiticity of the generating set must be preserved under the action of $\mathcal{Q}$. Hence we can construct a random Clifford element $\mathcal{Q}$ by inductively constructing a random symplectic matrix C and vector h.

Since h corresponds to keeping track of negative signs, the binary entries of h can be chosen uniformly at random. C is inductively constructed column by column where the first n columns correspond to the images of X_1 through X_n, and the last n columns correspond to the images of Z_1 through Z_n (all of which are written in binary notation as in [41]). Preservation of commutation relations is phrased through the symplectic inner product and so at each step one chooses the new column by finding a random solution to a system of linear equations which represents the inner product conditions. Since randomly choosing $2n$ elements of the Pauli group that satisfy the required commutation relations is equivalent to inductively choosing random solutions to $2n$ sets of linear equations (and each set requires at most $O(n^3)$ operations to solve), we can produce a random Clifford element in $O(n^4)$ (classical) operations.

Solution to 3: Any Clifford element can be decomposed into a sequence of $O(n^2)$ one and two-qubit generators in $O(n^3)$ time [41] (alternatively, there are slower methods which produce a "canonical" decomposition into $O(n^2/\log n)$ generators [1]). The method for the decomposition utilizes the symplectic representation of the Clifford group discussed above, ie. every Clifford element $\mathcal{Q}$ is represented up to phase by a binary, symplectic matrix C and a binary vector h. The main goal is to decompose C into generators as the negative signs represented by h can be accounted for via multiplication by single-qubit Pauli operators. The main theorem used in the decomposition of Clifford elements is theorem 4 of [41] which states that if C is a binary symplectic matrix then C can be decomposed as a product of five binary symplectic matrices, which we denote by T_1 through T_5.

These five symplectic matrices can be decomposed into symplectic matrices representing 1 and 2-qubit Clifford operations that correspond to Hadamard's, single qubit $\frac{\pi}{2}$-rotations about σ_Z, two-qubit $\frac{\pi}{2}$-rotations about $\sigma_Z \otimes \sigma_Z$, two-qubit permutation operations and $CNOT$ operations. This can be condensed into the following result:

Every Clifford operation $\mathcal{Q}$ can be realized by a sequence of one and two-qubit Clifford operations which consists of the following six rounds of operations:

1. An initial round of single-qubit Pauli operators,

2. Applying a sequence of $CNOT$ and two-qubit permutation operations,

3. Applying a sequence of $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$ followed by a sequence of $\frac{\pi}{2}$ rotations about σ_Z,

4. Applying a sequence of Hadamard operations,

5. Applying a sequence of $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$ followed by a sequence of $\frac{\pi}{2}$ rotations about σ_Z,

6. Applying a final round of $CNOT$ and two-qubit permutation operations.

Note that for rounds 3, 4 and 5 the operations in each round commute and can be performed in any order.

The time-complexity in decomposing a symplectic matrix into the sequence of one and two-qubit Clifford operations given above is $O(n^3)$ since one needs to solve linear systems of equations to obtain T_1 through T_5. In many cases one would like to have a decomposition of a Clifford element into a particular generating set for the Clifford group, such as $G_n := \{H, S, CNOT\}$ which consists of Hadamard's (H) and phase gates (S) on each qubit, as well as $CNOT$ gates on all pairs of qubits. There are $n^2 + n$ elements in G_n and it is a straightforward process to decompose the operations in 1 through 6 above into H, S and $CNOT$ gates.

Solution to 4: The $m+1$'th Clifford element is deterministically chosen such that composing it with the first m elements maps the initial state back to itself. By the Gottesman-Knill theorem we can keep track of the state of the system over the $mO\left(n^2\right)$ generating elements obtained from the m decompositions using $O\left(mn^4\right)$ classical operations. We are then left with a stabilizer state that needs to be mapped back to the original state via a Clifford operation. This is a straightforward process as we need only create a Clifford operation that maps the n generators of the output stabilizer state to the n generators of the input stabilizer state. There remains some freedom in the construction of the Clifford element since every Clifford element is uniquely determined by its action on $2n$ generators of the Pauli group satisfying the correct commutation relations, however the total time-cost in "mocking" up this Clifford gate is no more than $O\left(n^4\right)$ classical operations. Thus the $m+1$'th gate can be constructed and decomposed into generators using $O\left(mn^4\right) + O\left(n^3\right) = O\left(mn^4\right) = mO\left(n^4\right)$ classical operations.

In total, for an n-qubit system, we can efficiently choose Clifford gates uniformly at random and decompose each gate into a canonical subsequence of $O\left(n^2\right)$ elements from the

generating set G_n. The total time complexity in choosing, decomposing and implementing a random Clifford is $O\left(n^4\right) + O\left(n^3\right) + O\left(n^2\right) = O\left(n^4\right)$. The number of trials K one needs to perform to estimate $\mathcal{F}_g(m, \psi)$ to an accuracy ϵ with probability at least $1 - \delta$ is given by Eq. (2.110) which is independent of m and n. Thus for each m the time complexity of choosing m random Clifford elements, decomposing each into a sequence of generators and implementing the sequence is $mO\left(n^4\right)$. The inverse gate requires $O\left(n^4\right)$ operations to find, decompose and implement. Hence each sequence requires $(m+1)O\left(n^4\right)$ operations. Averaging over $K = \frac{\ln\left(\frac{2}{\delta}\right)(b-a)^2}{2\epsilon^2}$ trials implies for each m a total of $K(m+1)O\left(n^4\right)$ operations need to be performed. If M is the maximum sequence length then the total time complexity is no more than

$$\begin{aligned} \sum_{m=1}^{M-1} K(m+1)O\left(n^4\right) &= KO\left(n^4\right) \sum_{i=2}^{M} i \\ &= KO\left(n^4\right) \left(\frac{(M+2)(M-1)}{2}\right). \end{aligned} \tag{2.112}$$

Setting $[a, b] = [0, 1]$ implies a total time-complexity no larger than,

$$\frac{O\left(n^4\right)(M+2)(M-1)\ln(2/\delta)}{4\epsilon^2} \tag{2.113}$$

which implies the protocol is scalable in n.

2.7 Numerical Examples

As an example of the procedure, we first consider the case of benchmarking a single qubit under time-independent unitary errors with no state-preparation or measurement errors. For each $\mathcal{C}_j$, the unitary error was constructed by first finding the Hamiltonian that generates the Clifford operation via $\mathcal{C}_j(\rho) = \exp(-iH_j)\rho\exp(iH_j)$. For each H_j, the unitary $\exp(-iH_j)$ was diagonalized,

$$\exp(-iH_j) \sim \begin{bmatrix} e^{-i\lambda_j^1} & 0 \\ 0 & e^{-i\lambda_j^2} \end{bmatrix}$$

and to simulate the error, the eigenvalues were perturbed by a factor of δ to obtain,

$$\begin{bmatrix} e^{-i\lambda_j^1(1+\delta)} & 0 \\ 0 & e^{-i\lambda_j^2(1+\delta)} \end{bmatrix}.$$

Physically this corresponds to over/under rotations around H_j.

Two cases for δ were analyzed: $\delta = 0.1$ (case A) and δ chosen uniformly at random in the range $[0.075, 1.125]$ (case B). Numerical values for $F_{\text{seq}}(m, \psi)$ are shown in Fig. 2.2 as blue points. Note that we have subtracted the constant offset in the model so that pure exponentials appear as straight lines on the semi-log plot. In the present case $B_1 = B_0 = 1/2$ since the noise is unital and there are no state preparation or measurement errors. For both cases the first order model (green line) represents the data extremely well while the zeroth order (red dashed) only approximates the sequence fidelity when the variation in δ is small (case A). Furthermore, in case B the non-exponential behaviour of the average sequence fidelity is visible. This is also apparent in Table 2.3 which shows that the gate-dependence fit parameter $q - p^2$ is much larger for case B than case A.

We also considered two other error models of practical relevance: unitary error with depolarizing noise and unitary error with amplitude damping. The depolarizing and damping parameters for each Clifford element were chosen randomly in 0.9875 ± 0.01 with the unitary error chosen in the same way as case A. The results are summarized in Table 2.3 - in both these cases the simulations are well approximated by the zeroth order solution. These results illustrate that the zeroth-order randomized benchmarking model gives a robust estimate of the error-rate for a variety of realistic error models

In addition, for the two cases of pure unitary noise we have fit the numerical data to both the first order formula given in Eq. (2.54) (which we call r_1) as well as the formula given in Eq. (1) of [24] (which we call r_2) in order to obtain estimates of the average error rate r. We have chosen to compare to Eq. (1) of Ref. [24] because they explicitly assume unital noise and no measurement errors. Their model is given by

$$\mathcal{F}_g(m, \psi) = \frac{1}{2} + \frac{1}{2}(1 - d_{if})(1 - 2E_g)^m \tag{2.114}$$

where E_g is a parameter called the "error per gate" and d_{if} is a parameter that is claimed to represent state-preparation and measurement errors. This model is close to our zeroth order model given by Eq. (2.34). Indeed, it is straightforward to show that for a single-qubit,

$$p = 1 - 2r \tag{2.115}$$

so r can be identified with E_g. When the noise is unital with no measurement errors, $B_0 = \frac{1}{2}$ which is equal to their constant offset. Now, if one writes $A_0 = \frac{1}{2}(1 - d_{if})$ under the assumptions of unital noise and no measurement errors then it turns out,

$$d_{if} = 2(1 - \mathrm{tr}\left(|\psi\rangle\langle\psi|\Lambda(\rho_\psi)\right)). \tag{2.116}$$

Hence under their model this parameter represents their state-preparation errors.

The results are included in Table 2.2 along with the relative error when compared with the true average error rate r. In both cases fitting to the first order formula Eq. (2.54) (ie. r_1) produces a better fit value for the average error rate. Note that we have written the error values to two decimal places. Thus, while it appears r_1 and r_2 are the same in the case of Unitary A, they are actually different, which leads to different values of the relative error in r_1 and r_2.

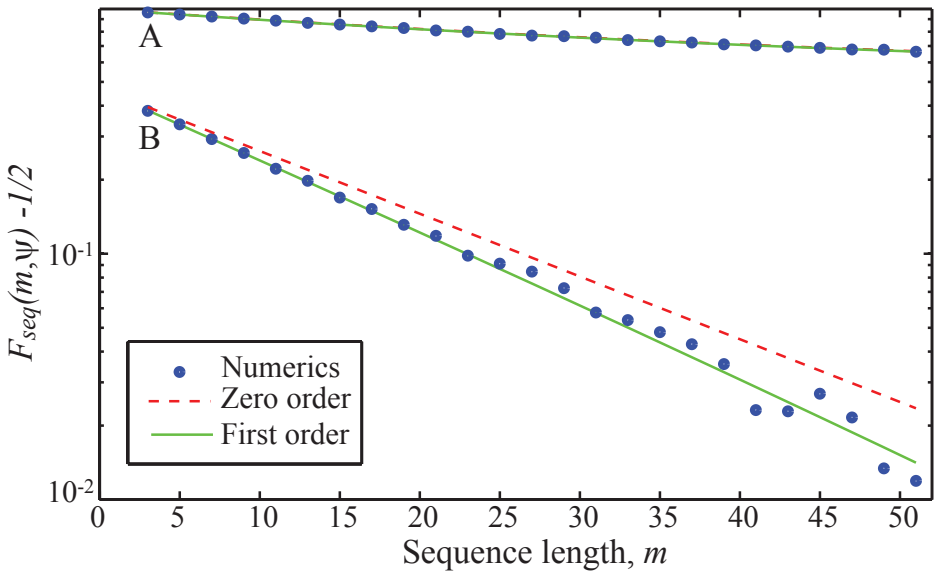

Figure 2.2: Average sequence fidelity as a function of sequence length for error models consisting of purely unitary noise; the first order model is required in case B where there is larger variation in the noise.

	Unitary A	Unitary B	Unitary and Dep.	Unitary and T_1
p	0.980	0.943	0.982	0.988
r	1.05e-2	2.85e-2	8.75e-3	5.85e-3
$q - p^2$	-2.73e-4	-6.83e-3	-2.77e-8	-2.80e-8

Table 2.1: Numerical results for the parameter p, error rate r and gate-dependence measure $q - p^2$ for the four cases of noise models considered; see text for details.

	Unitary A	Unitary B
r	1.05e-2	2.85e-2
r_1	1.05e-2	2.95e-2
relative error in r_1	0.569%	3.36%
r_2	1.05e-2	3.16e-2
relative error in r_2	0.729%	10.9%

Table 2.2: Fit values and relative errors of average error rate r for unitary noise; our first order model provides a more accurate estimate of the error-rate r than that of Ref. [24].

2.8 Conclusion

We have shown that randomized benchmarking provides a scalable method for benchmarking the set of Clifford gates. The protocol allows for time and gate-dependent noise and the fitting models for the fidelity function take into account state preparation and measurement errors. In addition to providing an estimate of the average fidelity across all Clifford gates, the first order model provides a measure of the gate-dependence of the noise.

We have provided here rigorous proofs of both the conditions for the validity of the protocol, as well as the scalability of the protocol in the number of qubits n comprising

the system. We have also established an exact relationship between the average fidelity estimate provided by the protocol and a stronger characterization of the average error operator strength given by the diamond norm for the case of when the average error operator over all gates is equal to a Pauli channel. The proof of this relationship utilizes a semidefinite program for computing the diamond norm [144] which has the potential to establish further connections between these two notions of error strength.

2.9 Discussion

One of the main questions arising from our protocol is whether one can benchmark non-Clifford gates. While benchmarking the full unitary group would be ideal, this is a provably inefficient task since tracking the state of the system through a sequence composed of Clifford operations and just one non-Clifford unitary is inefficient in n. On the other hand as we have shown here benchmarking the Clifford group is an efficient task. There are various reasons why benchmarking the Clifford group provides significant information for both fault-tolerant quantum computation as well as obtaining a benchmark for a generating set of the full unitary group. First, any realistic implementation of a quantum computer will have to take advantage of error-correcting codes in order to perform fault-tolerant quantum computation. The fact that most of the codes used in fault-tolerance theory are stabilizer codes implies that the encoding and decoding operations that have to be performed can be chosen to be Clifford operations. Moreover, depending on the form of the stabilizer code, it is likely that the majority of fundamental gates used in the computation are Clifford gates. For instance the seven qubit code [27, 135] has the property that encoded Clifford gates are applied transversally and so the encoded Clifford gates are comprised of Clifford gates on the physical qubits. Hence in such cases, a benchmark of the fundamental Clifford operations provides direct information regarding the robustness of encoding/decoding schemes and potentially the fidelity of the overall computation.

Second, the unitary group can be generated by adding just one single-qubit rotation not in the Clifford group (for instance the $\frac{\pi}{8}$-gate). Hence a benchmark for the Clifford group can provide useful information regarding a benchmark for a generating set of the full unitary group. Lastly, there is a model of quantum computation in which universal computation can be performed via Clifford gates, preparation of single-qubit non-stabilizer ancilla states called magic states [21] and measurements in the computational basis. Hence in this model of quantum computation the only physical gates that need to be benchmarked for universal quantum computation are Clifford gates.

Various other interesting questions and comments arise from the benchmarking analysis

presented here. First, there is a key point to emphasize regarding the zeroth and first order fitting models. As depicted in [99] there exist physically relevant noise models for which when the true value of the depolarization fidelity parameter p is used, the first order model fits the experimental data much better than the zeroth order model. However, it may be the case that a least squares fitting procedure using the zeroth order model produces a very good fit to the experimental data, albeit an incorrect value for p. Therefore in order to obtain a more accurate value for p one should always use the first order fitting model unless prior knowledge of the noise indicates that it is effectively gate-independent.

It will be useful to obtain a better understanding for when a least squares fitting procedure using the zeroth order model produces a value for p that is close to its true value. Clearly in the gate-independent case the zeroth order model fits the fidelity decay curve exactly. Moreover for weakly gate-dependent noise one can see from our continuity argument that the zeroth order model is still a sufficient fitting function for the fidelity decay curve. Hence the most interesting case to analyze is when there is a non-negligible amount of gate-dependence in the noise and the condition for using the first order model to fit the decay curve is satisfied. A useful test that would indicate gate-dependence in the noise, and thus the validity of the value of p obtained from fitting to the zeroth order model, is to perform the least squares fitting procedure using both the zeroth and first order fitting models. If the estimates of p obtained in each case differ significantly then the zeroth order model must be a poor choice of fitting function even though it may fit the data well. In this case the noise must have a strong gate-dependence because otherwise $q - p^2$ would be small which implies the two fitting functions would produce similar estimates for p.

An interesting question is how to extract a meaningful average error rate over a generating set of the Clifford group, for instance G_n defined previously, from the average error rate r over the entire Clifford group. One might argue that benchmarking a generating set for the Clifford group is sufficient for benchmarking the full Clifford group, however it is entirely plausible that noise correlations between the n physical qubits creates large errors on elements of Clif_n, even when the errors on the generating set can be controlled [103]. In fact an assumption that is often made in fault-tolerant estimates is that the correlation in noise between qubits is either small or can be ignored.

With regards to scalability, while we have shown the protocol itself is scalable in n, a useful direction for further research would be an analysis of how the sufficient condition of weak average variation of the noise depends on n. Since multi-qubit gates are realized by implementing a sequence of generators, the noise associated to a multi-qubit Clifford element is given by the noise associated to the entire sequence of generators. A determination of whether these noise operators continue to satisfy the sufficient condition when it

is met for small numbers of qubits will be useful for understanding the applicability of the protocol.

Rigorous fault-tolerant analyses sometimes invoke the diamond norm as a measure of the error strength rather than the weaker characterization provided by the average fidelity. Hence it is desirable to find relationships between these two quantities that is more general than the special case of random Pauli errors presented here. As mentioned above, the semidefinite program we have used to deduce the relationship appears to be a promising tool for further research in this area. From the expression given in Eq. (A.43) one can see that the diamond norm is essentially a "worst-case" maximization over input (entangled) states. In quantum computation it is the case that the measure of accessible states (states that can be reached in polynomial time using a generating set for the unitary group) is equal to 0. Hence there is a high probability that the maximization criteria demanded by the diamond norm is a much stronger condition than necessary for understanding the strength of the errors affecting the computation. This point becomes even more relevant for models of quantum computation such as adiabatic quantum computing where the computational model is more "algorithm-specific". In such cases there will be an even tighter restriction on the set of accessible states. An interesting direction of further research is to provide precise conditions for when the average fidelity provides an indication or bound on the error strength in terms of stronger characterizations such as the diamond norm.

Additionally, if one were able to obtain an estimate of the minimum gate fidelity from knowledge of the average fidelity they could use the direct relationship between the minimum gate fidelity and diamond norm given by Eq. (A.58) to obtain information about the error strength in terms of the diamond norm. A result that may be useful in this direction of research is the 'concentration of measure effect of the gate fidelity which implies that as n increases, the measure of the set of states which produce a fidelity close to the minimum yet far from the average is exponentially small in n (see Sec. D.2 or [98, 97]).

Lastly, we gather many of the randomized benchmarking results that have been obtained to date in various architectures:

The above results were obtained using the protocol of [84] since the experiments were carried out prior to the formulation of our protocol and in particular our first order fitting model for the fidelity decay.

	Error rate r
NMR	1.3×10^{-4} [124]
Ion traps	4.82×10^{-3} [84]
	8×10^{-4} [15]
	2×10^{-5} [24]
Optical lattices	1.4×10^{-4} [110]
Superconducting qubits	1.1×10^{-2} [32]
	7×10^{-3} [31]

Table 2.3: Results for the average error rate r from randomized benchmarking in various architectures for quantum computation.

Chapter 3

Properties of Quantum Channels and the Quantum Gate Fidelity

In certain models of computation, such as the standard circuit model, quantum information is ideally evolved by unitary operations. Experimentally, a unitary transformation $\mathcal{U}$ will not be performed perfectly and the actual (implemented) transformation is some general, and likely unknown, quantum operation $\mathcal{E}$. A natural question to ask is how distinguishable are $\mathcal{U}$ and $\mathcal{E}$ under an appropriate distance measure on quantum channels. The distinguishability of quantum operations has been well-studied in the literature [51, 78, 3, 56]. A comprehensive discussion of various types of distance measures on quantum channels along with an exhaustive set of criteria a useful distance measure should satisfy is given in [56].

One measure that is particularly useful in experimental protocols is the quantum gate fidelity. It can be obtained from the quantum channel fidelity which is a natural extension of the fidelity between quantum states to quantum channels (see Sec. A.4.2). The channel fidelity between two quantum operations $\mathcal{E}_1$ and $\mathcal{E}_2$ is the real-valued function on quantum states given by

$$\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}(\rho) = \left(\mathrm{tr}\sqrt{\sqrt{\mathcal{E}_1(\rho)}\mathcal{E}_2(\rho)\sqrt{\mathcal{E}_1(\rho)}}\right)^2$$

where ρ is an arbitrary mixed quantum state. When the input states are restricted to be pure and the two operations are a unitary $\mathcal{U}$ and a quantum operation $\mathcal{E}$, the above function is called the quantum gate fidelity and can be written as,

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(|\phi\rangle) = \mathrm{tr}\left(\mathcal{U}(|\phi\rangle\langle\phi|)\mathcal{E}(|\phi\rangle\langle\phi|)\right)$$

for pure states $|\phi\rangle$. The state-dependence of the gate fidelity can be removed by averaging over input states to obtain the *average gate fidelity*, or taking the minimum over all states which gives the *minimum gate fidelity*. These two distance measures satisfy some of the criteria in [56] to be a useful distance measure and, as seen in the previous chapter, play an important role in the experimental characterization of quantum processes.

Recently, methods have been developed for calculating exact expressions of both the average and minimum gate fidelity [56, 107, 46, 113, 98, 87] given a theoretical description of $\mathcal{U}$ and $\mathcal{E}$. As previously mentioned, an experimental procedure for obtaining a description of $\mathcal{E}$ is given by quantum process tomography [117, 33] which scales exponentially in the number of qubits n. As a result, there has recently been interest in providing efficient experimental procedures for characterizing certain features of $\mathcal{E}$ [40, 47, 134, 11, 119], such as the average gate fidelity. The randomized benchmarking protocol of the previous chapter is an example of such an experimental procedure where the average fidelity over a set of computational gates is estimated.

The average fidelity provides no information about *fluctuations* in the gate fidelity – i.e., how the error varies over input states. The variance is a useful diagnostic in terms of noise characterization as it can provide information about the minimum gate fidelity, which is relevant for fault-tolerant design. Moreover, large variance relative to a small average error could suggest that contributions to the average error are dominated by only a few very error-prone states. Addressing those states would then produce dramatic improvements in average fidelity. Large fluctuations may also indicate hidden high-fidelity information-preserving structures such as pointer bases or (approximate) decoherence-free subspaces and noiseless subsystems [17].

This chapter provides a detailed analysis of various properties of the gate fidelity and is based on the content of Ref.'s [97, 98]. First, in Sec. 3.1 we show that two distinct quantum channels can produce the same gate fidelity function. Specifically, if $d \geq 4$ then for any unitary operator $\mathcal{U}$ and full-rank quantum channel $\mathcal{E}_1$ there exists a quantum operation $\mathcal{E}_2$ (not equal to either of $\mathcal{E}_1$ or $\mathcal{E}_1^\dagger$) which satisfies $\mathcal{F}_{\mathcal{E}_1,\mathcal{U}}(|\psi\rangle) = \mathcal{F}_{\mathcal{E}_2,\mathcal{U}}(|\psi\rangle)$ for every pure state $|\psi\rangle$. Since depolarizing channels are full-rank, a corollary of this result is that if $d \geq 4$ there exist *non-depolarizing* channels $\mathcal{E}$ such that $\mathcal{F}_{\mathcal{E},\mathcal{I}}$ is constant on the set of pure states.

Next, in Sec. 3.2 and 3.3 we introduce a novel general method for calculating any moment of the gate fidelity, and apply it to calculate expressions for both the average and variance. Then in Sec. 3.4 and 3.5 we calculate an explicit expression for the single-qubit

case as well as an explicit upper-bound that holds for any dimension. The upper-bound depends only on the dimension of the system which allows us to deduce the asymptotic behaviour of the variance.

Sec. 3.6.1 uses Levy's lemma to calculate an upper bound on the probability that a randomly chosen state will produce a gate fidelity value that is far from the average. The measure of the deviating set of states converges to 0 exponentially quickly in the dimension of the quantum system. Section 3.6.2 shows how one can use these results to obtain alternative upper bounds for the variance of the gate fidelity than those found in Sec. 3.5. Some of the abstract statistical results obtained to this point are explained for the amplitude damping channel in Sec. 3.6.3. Sec. 3.6.4 ties many of these statistical results together by formalizing the notion of convergence to depolarization of quantum channels with respect to the gate fidelity. Lastly, Sec. 3.6.5 provides two methods for estimating the minimum gate fidelity using these statistical results and we conclude with a discussion of our results as well as future research directions in Sec.'s 3.7 and 3.8.

3.1 Non-Uniqueness of the Gate Fidelity

As mentioned in the introduction, the gate fidelity is particularly important in experimental quantum computation because the ideal transformation is a unitary superoperator, while the implemented (real) transformation is some general quantum operation. A question that arises is, if the intended unitary operation is $\mathcal{U}$, then does the gate fidelity on $\mathbb{CP}^{d-1}$ uniquely characterize the implemented quantum operation? Equivalently, if the unitary operator $\mathcal{U}$ is fixed then can there exist two distinct quantum channels $\mathcal{Q}$ and $\mathcal{R}$ satisfying $\mathcal{F}_{\mathcal{Q},\mathcal{U}} = \mathcal{F}_{\mathcal{R},\mathcal{U}}$? From Eq. (A.75) this question is equivalent to the problem of determining whether there exist two distinct quantum channels, that we continue to denote as $\mathcal{Q}$ and $\mathcal{R}$, such that $\mathcal{F}_{\mathcal{Q},\mathcal{I}} = \mathcal{F}_{\mathcal{R},\mathcal{I}}$.

It is clear that the gate fidelity is not unique in general by noting that if $\mathcal{E}$ is a channel such that $\mathcal{E} \neq \mathcal{E}^\dagger$ then

$$\operatorname{tr}\left(\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle\langle\psi|\right) = \operatorname{tr}\left(|\psi\rangle\langle\psi|\mathcal{E}^\dagger(|\psi\rangle\langle\psi|)\right).$$

The main theorem of this section, Theorem 2, shows that if $d \geq 4$ and $\mathcal{Q}$ is a full-rank quantum operation then there exists a quantum channel $\mathcal{R} \neq \mathcal{Q}^\dagger$ which produces the *same* gate fidelity function. In this context, full-rank means that the minimum number of Kraus operators required for $\mathcal{Q}$ is d^2. From Sec. A.2 this requirement is equivalent to the Choi

matrix of $\mathcal{Q}$ being positive-definite. A weaker condition under which the theorem still holds is discussed immediately after the proof of the theorem.

Theorem 2. *Suppose that dim($\mathcal{H}$)= $d \geq 4$ and $\mathcal{Q}$ is a quantum operation on $L(\mathcal{H})$ with a positive-definite Choi matrix. Then there exists a quantum channel $\mathcal{R} \neq \mathcal{Q}^\dagger$ (and $\mathcal{R} \neq \mathcal{Q}$) such that*

$$\mathcal{F}_{\mathcal{Q},\mathcal{I}} = \mathcal{F}_{\mathcal{R},\mathcal{I}}.$$

In order to prove Theorem 2 we will need the following lemma:

Lemma 1. *A linear superoperator Λ acting on $L(\mathcal{H})$ can be written as the difference between two quantum operations Λ_1 and Λ_2 satisfying $\mathcal{F}_{\Lambda_1,\mathcal{I}} = \mathcal{F}_{\Lambda_2,\mathcal{I}}$ if the following conditions are satisfied,*

1. $C(\Lambda)$ is the difference between two positive semi-definite operators A and B such that $tr_{\mathcal{H}_1} A = tr_{\mathcal{H}_1} B = \mathbb{1}$,

2. $(\mathcal{I} \otimes T)(C(\Lambda))$ has support on the anti-symmetric subspace of $\mathcal{H}_1 \otimes \mathcal{H}_2$ where $\mathcal{I} \otimes T$ represents the partial transpose operation on $L(\mathcal{H}_1 \otimes \mathcal{H}_2)$ (see Sec. A.2).

Proof. (Lemma)

First, suppose that $C(\Lambda)$ is equal to $A - B$ where A and B are positive semi-definite operators and $\mathrm{tr}_{\mathcal{H}_1} A = \mathrm{tr}_{\mathcal{H}_1} B = \mathbb{1}$. From Sec. A.2 these assumptions on A and B are equivalent to $A = C(\Lambda_1)$ and $B = C(\Lambda_2)$ for quantum operations Λ_1 and Λ_2. Thus by linearity, condition 1 is equivalent to $\Lambda = \Lambda_1 - \Lambda_2$ where Λ_1 and Λ_2 are quantum operations. Hence it remains to show that the second condition implies $\mathcal{F}_{\Lambda_1,\mathcal{I}} = \mathcal{F}_{\Lambda_2,\mathcal{I}}$.

Since the vec correspondence (see Sec. A.2) between $L(\mathcal{H})$ with the Hilbert-Schmidt inner product and $\mathcal{H}_1 \otimes \mathcal{H}_2$ with the standard inner product is an inner-product space isomorphism, for any A, B in $L(\mathcal{H})$,

$$\langle A, B \rangle = \mathrm{tr}\left(A^\dagger B\right) = \mathrm{vec}(A)^\dagger \mathrm{vec}(B) = \langle \mathrm{vec}(A), \mathrm{vec}(B) \rangle.$$

If $C(\Lambda)$ has spectral decomposition,

$$C(\Lambda) = \sum_i \lambda_i \mathrm{vec}(A_i)\mathrm{vec}(A_i)^\dagger,$$

then,

$$\begin{aligned}
\langle C(\Lambda), |m\rangle \otimes |n\rangle\langle k| \otimes \langle l| \rangle &= \mathrm{tr}\left[\left(\sum_i \lambda_i \mathrm{vec}(A_i)\mathrm{vec}(A_i)^\dagger\right)(|m\rangle \otimes |n\rangle\langle k| \otimes \langle l|)\right] \\
&= \sum_i \lambda_i \mathrm{vec}(A_i)^\dagger |m\rangle \otimes |n\rangle \mathrm{tr}\left[\mathrm{vec}\left(A_i\right)\langle k| \otimes \langle l|\right] \\
&= \sum_i \lambda_i \langle |k\rangle \otimes |l\rangle, \mathrm{vec}(A_i)\rangle\langle \mathrm{vec}(A_i), |m\rangle \otimes |n\rangle\rangle.
\end{aligned}$$

The vec correspondence again gives,

$$\begin{aligned}
\sum_i \lambda_i \langle |k\rangle \otimes |l\rangle, \mathrm{vec}(A_i)\rangle\langle \mathrm{vec}(A_i), |m\rangle \otimes |n\rangle\rangle &= \sum_i \lambda_i \mathrm{tr}\left(\left(|k\rangle\langle l|\right)^\dagger A_i\right) \mathrm{tr}\left(A_i^\dagger |m\rangle\langle n|\right) \\
&= \sum_i \lambda_i \langle k|A_i|l\rangle\langle n|A_i^\dagger|m\rangle \\
&= \mathrm{tr}\left(\Lambda\left(|l\rangle\langle n|\right)|m\rangle\langle k|\right)
\end{aligned}$$

and so,

$$\langle C(\Lambda), |m\rangle \otimes |n\rangle\langle k| \otimes \langle l| \rangle = \mathrm{tr}\left(\Lambda\left(|l\rangle\langle n|\right)|m\rangle\langle k|\right). \tag{3.1}$$

Noting that,

$$\begin{aligned}
\langle C(\Lambda), |m\rangle \otimes |n\rangle\langle k| \otimes \langle l| \rangle &= \mathrm{tr}\left(C(\Lambda)^\dagger |m\rangle \otimes |n\rangle\langle k| \otimes \langle l|\right) \\
&= \mathrm{tr}\left(C(\Lambda)\left[|m\rangle\langle k| \otimes \left(|l\rangle\langle n|\right)^T\right]\right) \\
&= \mathrm{tr}\left(C(\Lambda)\left[\mathcal{I} \otimes T\left(|m\rangle\langle k| \otimes |l\rangle\langle n|\right)\right]\right)
\end{aligned} \tag{3.2}$$

and,

$$\mathrm{tr}\left(C(\Lambda)\left[\mathcal{I}\otimes T\left(|m\rangle\langle k|\otimes|l\rangle\langle n|\right)\right]\right)=\mathrm{tr}\left(\left[\mathcal{I}\otimes T\left(C(\Lambda)\right)\right]|m\rangle\langle k|\otimes|l\rangle\langle n|\right),$$

for any $|\psi\rangle\in\mathbb{CP}^{d-1}$,

$$\mathrm{tr}\left(\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle\langle\psi|\right)=\mathrm{tr}\left(\left[\mathcal{I}\otimes T\left(C(\Lambda)\right)\right]|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right).$$

Hence $\mathrm{tr}\left(\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle\langle\psi|\right)=0$ if and only if $\mathrm{tr}\left(\left[\mathcal{I}\otimes T\left(C(\Lambda)\right)\right]|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right)=0$.

In total, the above discussion shows that the conditions:

1. $C(\Lambda)$ is the difference between two positive semi-definite operators A and B such that $\mathrm{tr}_{\mathcal{H}_1}A=\mathrm{tr}_{\mathcal{H}_1}B=\mathbb{1}$,

2. For every $|\psi\rangle\langle\psi|$, $\mathrm{tr}\left(\left[\mathcal{I}\otimes T\left(C(\Lambda)\right)\right]|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right)=0$,

are satisfied if and only if Λ can be written as the difference between two quantum operations Λ_1 and Λ_2 satisfying $\mathcal{F}_{\Lambda_1,\mathcal{I}}=\mathcal{F}_{\Lambda_2,\mathcal{I}}$.

Let the symmetric and anti-symmetric subspace in $\mathcal{H}_1\otimes\mathcal{H}_2$ be denoted sym(2,d) and a-sym(2,d) respectively so that $\mathcal{H}_1\otimes\mathcal{H}_2=\mathrm{sym}(2,d)\oplus\text{a-sym}(2,d)$. Since every state $|\psi\rangle$ satisfies $|\psi\rangle\otimes|\psi\rangle\in\mathrm{sym}(2,d)$, if $\left(\mathcal{I}\otimes T\right)\left(C(\Lambda)\right)$ has support on a-sym$(2,d)$ then

$$\mathrm{tr}\left(\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle\langle\psi|\right)=\mathrm{tr}\left(\left[\left(\mathcal{I}\otimes T\right)\left(C(\Lambda)\right)\right]|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right)=0$$

for every $|\psi\rangle$. Thus the conditions:

1. $C(\Lambda)$ is the difference between two positive semi-definite operators A and B such that $\mathrm{tr}_{\mathcal{H}_1}A=\mathrm{tr}_{\mathcal{H}_1}B=\mathbb{1}$,

2. $\left(\mathcal{I}\otimes T\right)\left(C(\Lambda)\right)$ has support on a-sym(2,d),

are sufficient for Λ to be the difference between two quantum operations which produce the same gate-fidelity.

□

Theorem 2 can now be proven using Lemma 1.

Proof. (Theorem)

First, let $d = 4$ so that $\mathcal{H}_1$ and $\mathcal{H}_2$ as defined above are both identified with $\mathbb{C}^4$, and suppose $\mathcal{Q}$ is such that $C(\mathcal{Q}) > 0$. $\mathcal{R}$ is explicitly constructed by first showing that there is an element of $L(\mathcal{H}_1 \otimes \mathcal{H}_2)$ satisfying the two conditions from Lemma 1. Define,

$$\begin{aligned}
|\alpha_1\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \;\; |\beta_1\rangle = \frac{1}{\sqrt{2}}(|23\rangle - |32\rangle), \\
|\alpha_2\rangle &= \frac{1}{\sqrt{2}}(|02\rangle - |20\rangle), \;\; |\beta_2\rangle = \frac{1}{\sqrt{2}}(|13\rangle - |31\rangle), \\
|\alpha_3\rangle &= \frac{1}{\sqrt{2}}(|03\rangle - |30\rangle), \;\; |\beta_3\rangle = \frac{1}{\sqrt{2}}(|12\rangle - |21\rangle).
\end{aligned}$$

These six vectors form an orthonormal basis for a-sym(2,4). Define $G \in L(\mathcal{H}_1 \otimes \mathcal{H}_2)$ via the equation,

$$(\mathcal{I} \otimes T)(G) = |\alpha_1\rangle\langle\beta_1| + |\alpha_2\rangle\langle\beta_2| + |\alpha_3\rangle\langle\beta_3| + |\beta_1\rangle\langle\alpha_1| + |\beta_2\rangle\langle\alpha_2| + |\beta_3\rangle\langle\alpha_3|.$$

It is straightforward to verify that G is Hermitian (ie. G is Hermitian if and only if $\mathcal{I} \otimes T(\mathcal{G})$ is Hermitian), $\mathrm{tr}_{\mathcal{H}_1}(G) = \mathrm{tr}_{\mathcal{H}_1}((\mathcal{I} \otimes T)(G)) = 0$ and $(\mathcal{I} \otimes T)(G)$ has support on a-sym$(2, 4)$.

Let $\mathcal{G}$ be the unique linear superoperator such that $C(\mathcal{G}) = G$. Since $C(\mathcal{Q}) > 0$ there exists $\epsilon > 0$ depending on both $\mathcal{Q}$ and $\mathcal{G}$ such that

$$C(\mathcal{Q}) + \epsilon C(\mathcal{G}) \geq 0.$$

Thus $\epsilon\mathcal{G}$ is such that,

1. $C(\epsilon\mathcal{G}) = C(\mathcal{Q} + \epsilon\mathcal{G}) - C(\mathcal{Q})$ with $C(\mathcal{Q})$, $C(\mathcal{Q} + \epsilon\mathcal{G}) \geq 0$ and $\mathrm{tr}_{\mathcal{H}_1} C(\mathcal{Q} + \epsilon\mathcal{G}) = \mathrm{tr}_{\mathcal{H}_1} C(\mathcal{Q}) = \mathbb{1}$,

2. $(\mathcal{I} \otimes T)(C(\epsilon\mathcal{G})) = \epsilon\,(\mathcal{I} \otimes T)(G)$ has support on a-sym(2,d).

Hence from Lemma 1, $\mathcal{Q}$ and $\mathcal{R} := \mathcal{Q} + \epsilon\mathcal{G}$ are two quantum operations that produce the same gate fidelity. Up to finding an explicit value for ϵ this proves the theorem for $d = 4$.

To find a value for ϵ note that since $C(\mathcal{Q}) > 0$, the smallest eigenvalue of $C(\mathcal{Q})$, denoted $\lambda_{\min}^{\mathcal{Q}}$, is strictly greater than 0. Therefore for every vector $|\phi\rangle \in \mathbb{C}^4 \otimes \mathbb{C}^4$,

$$\langle\phi|C(\mathcal{Q})|\phi\rangle \in \left[\lambda_{\min}^{\mathcal{Q}}, \|C(\mathcal{Q})\|_\infty\right]$$

Moreover, since $\langle\phi|\epsilon C(\mathcal{G})|\phi\rangle \in [-\epsilon\|C(\mathcal{G})\|_\infty, \epsilon\|C(\mathcal{G})\|_\infty]$,

$$\langle\phi|C(\mathcal{Q} + \epsilon\mathcal{G})|\phi\rangle \in \left[\lambda_{\min}^{\mathcal{Q}} - \epsilon\|C(\mathcal{G})\|_\infty, \|C(\mathcal{Q})\|_\infty + \epsilon\|C(\mathcal{G})\|_\infty\right].$$

Therefore in order for $C(\mathcal{Q} + \epsilon\mathcal{G}) \geq 0$ to be satisfied it must be that

$$0 < \epsilon \leq \frac{\lambda_{\min}^{\mathcal{Q}}}{\|C(\mathcal{G})\|_\infty}.$$

Lastly, suppose $d > 4$. Since the vector space spanned by $\{|\alpha_1\rangle, |\alpha_2\rangle, |\alpha_3\rangle, |\beta_1\rangle, |\beta_2\rangle, |\beta_3\rangle\}$ is a subspace of a-sym(2,d), $\mathcal{Q} + \epsilon\mathcal{G}$ can be defined in the same manner as above which proves the theorem.

□

Note that Theorem 2 still holds if the condition of $\mathcal{Q}$ being full-rank is relaxed to the existence of $\epsilon > 0$ such that $C(\mathcal{Q}) \geq C(-\epsilon\mathcal{G})$. However in this more general case it is difficult to obtain an explicit expression for ϵ under which the theorem holds.

The following corollary follows immediately from Theorem 2.

Corollary 2. *Let $dim(\mathcal{H}) = d \geq 4$. Suppose $\mathcal{Q}$ is a depolarizing channel on $L(\mathcal{H})$ of the form*

$$\mathcal{Q}(A) = pA + (1-p)tr(A)\frac{\mathbb{1}}{d}$$

where $p \in [0, 1)$ and let $\mathcal{G}$ be the linear superoperator from Theorem 2. Then for any $\epsilon \in \left(0, \frac{1-p}{d\|C(\mathcal{G})\|_\infty}\right]$, $\mathcal{R} = \mathcal{Q} + \epsilon\mathcal{G}$ is a non-depolarizing quantum operation with $\mathcal{F}_{\mathcal{Q},\mathcal{I}} = \mathcal{F}_{\mathcal{R},\mathcal{I}}$.

Proof. Since $\mathcal{Q}$ is depolarizing with $p \in [0,1)$, $C(\mathcal{Q})$ is a positive matrix. Thus Theorem 2 gives both the existence and construction of $\mathcal{R}$ in terms of $\mathcal{G}$. The fact that ϵ lies in $\left(0, \frac{1-p}{d\|C(\mathcal{G})\|_\infty}\right]$ follows from the fact that $\lambda_{\min}^{\mathcal{Q}} = \frac{1-p}{d}$. To see that $\lambda_{\min}^{\mathcal{Q}} = \frac{1-p}{d}$, note that since

$$\frac{1}{d^2}\sum_{i=0}^{d^2-1} P_i \rho P_i = \frac{\mathbb{1}}{d} \tag{3.3}$$

we have,

$$\begin{aligned} \mathcal{Q}(\rho) &= p\mathbb{1}\rho\mathbb{1} + \frac{1-p}{d^2}\sum_{i=0}^{d^2-1} P_i \rho P_i \\ &= \left(p + \frac{1-p}{d^2}\right)\mathbb{1}\rho\mathbb{1} + \frac{1-p}{d^2}\sum_{i=1}^{d^2-1} P_i \rho P_i. \end{aligned} \tag{3.4}$$

Hence the χ-matrix of $\mathcal{Q}$ with respect to the Pauli basis is diagonal, has $(0,0)$ element given by $p + \frac{1-p}{d^2}$, and has all other diagonal elements equal to $\frac{1-p}{d^2}$. Since this matrix is unitarily equivalent to $J(\mathcal{Q})$ (ie. it is equal to $J(\mathcal{Q})$ written in the generalized Bell-basis), these are in fact the eigenvalues of $J(\mathcal{Q})$. Hence, the eigenvalues of $C(\mathcal{Q})$ are $dp + \frac{1-p}{d}$ (multiplicity 1) and $\frac{1-p}{d}$ (multiplicity d^2-1) which implies

$$\lambda_{\min}^{\mathcal{Q}} = \frac{1-p}{d}. \tag{3.5}$$

□

Corollary 2 shows that the gate fidelity cannot always distinguish between depolarizing and non-depolarizing quantum channels. The following is a straightforward result of Eq. (A.77) and Corollary 2,

Corollary 3. *If $dim(\mathcal{H}) = d \geq 4$ then there exist non-depolarizing quantum channels $\mathcal{E}$ such that $\mathcal{F}_{\mathcal{E},\mathcal{I}}$ is constant on $\mathbb{CP}^{d-1}$.*

In terms of the Bloch representation of quantum states [16, 123], the action of a depolarizing channel is to isotropically shrink the Bloch object. Corollary 3 shows that even if

the gate fidelity between $\mathcal{E}$ and $\mathcal{I}$ is a constant function, one is unable to deduce whether $\mathcal{E}$ isotropically shrinks the Bloch object.

3.2 Calculating the Variance of the Gate Fidelity

Let us now turn to the problem of calculating the variance of $\mathcal{F}_{\mathcal{E},\mathcal{U}}$. For ease of notation, $\mathcal{F}_{\mathcal{E},\mathcal{U}}$ will be denoted simply by $\mathcal{F}$ for the rest of this section. We have,

$$\mathrm{Var}\left(\mathcal{F}\right) = \overline{\mathcal{F}^2} - \overline{\mathcal{F}}^2.$$

The existence of a basis of linear operators $\{P_a\}_{a=0}^{d^2-1}$ with the properties listed in Eq. (B.6) will play an important role in our derivation. Our ultimate goal is the expression Eq. (3.22), which is written in terms of the χ-matrix of $\Lambda \equiv \mathcal{U}^\dagger \circ \mathcal{E}$ that is obtained from expanding any set of Kraus operators for Λ in terms of $\{P_a\}_{a=0}^{d^2-1}$ (see Sec. A.2). Also, from Sec. A.2, the Jamiolkowski representation of Λ written in the basis $\{(P_a \otimes \mathbb{1})\,|\psi_0\rangle\}_{a=0}^{d^2-1}$ is equal to χ written with respect to $\{P_a\}_{a=0}^{d^2-1}$. Hence without loss of generality we sometimes refer to χ as the Jamiolkowski representation of Λ, however it is important to keep in mind what space χ acts upon, as well as what basis χ is written with respect to. We will rely heavily on the theory of Sec.'s A.2 and C.3 in this section and so the reader may want to look over these sections beforehand.

3.2.1 Average Gate Fidelity

To determine $\mathrm{Var}(\mathcal{F})$, we need to calculate both $\overline{\mathcal{F}}$ and $\overline{\mathcal{F}^2}$. Fortunately, the tools of Sec. C.3 can be used calculate *any* moment of $\mathcal{F}$, although the calculation of $\overline{\mathcal{F}^n}$ gets rapidly harder with increasing n. Hence we begin with $\overline{\mathcal{F}}$, which is already well-known [107], to give some intuition for our method of calculating any moment.

We begin by expanding the state-dependent gate fidelity in terms of Λ's Kraus operators $\{K_i\}$,

$$\begin{aligned}
\mathcal{F}(|\psi\rangle\langle\psi|) &= \mathrm{tr}\left[\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle\langle\psi|\right] \\
&= \sum_i \mathrm{tr}\left[K_i|\psi\rangle\langle\psi|\right]\mathrm{tr}\left[K_i^\dagger|\psi\rangle\langle\psi|\right] \\
&= \sum_i \mathrm{tr}\left[\left(K_i \otimes K_i^\dagger\right)|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|\right].
\end{aligned}$$

This expression is a Hilbert-Schmidt inner product between (i) a term including all the Kraus operators, and (ii) a term including all the $|\psi\rangle$-dependence. To average over ψ, we need only average the second term. Using Eq. (C.18),

$$\begin{aligned}\overline{\mathcal{F}} &= \int \mathcal{F}(|\psi\rangle\langle\psi|)d\psi \\ &= \sum_i \mathrm{tr}\left[\left(K_i \otimes K_i^\dagger\right)\overline{|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|}\right] \\ &= \sum_i \mathrm{tr}\left[\left(K_i \otimes K_i^\dagger\right)\frac{\pi_{\mathrm{sym}}(2,d)}{\mathrm{tr}\left(\pi_{\mathrm{sym}}(2,d)\right)}\right]\end{aligned}$$

where $\pi_{\mathrm{sym}}(2,d)$ is the projector onto the symmetric subspace of $\mathcal{H}^{\otimes 2} \equiv \left(\mathbb{C}^d\right)^{\otimes 2}$. We now expand $\pi_{\mathrm{sym}}(2,d)$ as a sum of permutation operators, invoke Eq. (C.20) to evaluate the traces, and use Eq. (C.19) to evaluate the normalization:

$$\begin{aligned}\overline{\mathcal{F}} &= \frac{1}{2\mathrm{tr}\left[\pi_{\mathrm{sym}}(2,d)\right]}\sum_i\sum_{\sigma\in S_2}\mathrm{tr}\left[\left(K_i \otimes K_i^\dagger\right)\mathcal{P}_\sigma\right] \\ &= \frac{\sum_i\left(\mathrm{tr}\left[K_i\right]\mathrm{tr}\left[K_i^\dagger\right]\right)+d}{d^2+d}.\end{aligned}$$

We can also write $\overline{\mathcal{F}}$ in terms of the Jamiolkowski representation χ of Λ. Since $\Lambda(\rho) = \sum_{l,m}\chi_{l,m}P_l\rho P_m$, the same calculation yields

$$\begin{aligned}\overline{\mathcal{F}} &= \frac{2}{d^2+d}\sum_{l,m}\mathrm{tr}\left[\chi_{l,m}\left(P_l \otimes P_m\right)\pi_{\mathrm{sym}}(2,d)\right] \\ &= \frac{1}{d^2+d}\sum_{l,m}\chi_{l,m}\left(\mathrm{tr}\left[P_l\right]\mathrm{tr}\left[P_m\right]+\mathrm{tr}\left[P_lP_m\right]\right) \\ &= \frac{\chi_{0,0}d+1}{d+1},\end{aligned}$$

which agrees with the results from Refs. [46, 107]. Recalling that $\chi_{0,0} = \mathrm{tr}\left[\chi\chi_0\right]$ we have,

$$\overline{\mathcal{F}} = \frac{\mathrm{tr}\left[\chi\chi_0\right]d+1}{d+1}. \tag{3.6}$$

We observe that $\mathrm{tr}\left[\chi\chi_0\right]$ represents the overlap of Λ with the identity channel, and therefore

how much Λ leaves the input state unchanged. It is also a unitary invariant of Λ; $\chi_{0,0}$ does not change if we rotate Λ by a unitary channel $\mathcal{U}$, mapping $\Lambda \to \mathcal{U}^{-1} \circ \Lambda \circ \mathcal{U}$.

3.2.2 Variance of the Gate Fidelity

Now, let's look at the calculation of $\overline{\mathcal{F}^2}$. As done previously, we expand $\mathcal{F}^2$ in terms of Λ's Kraus operators,

$$\begin{aligned}
\mathcal{F}^2(|\psi\rangle\langle\psi|) &= \mathrm{tr}\left[\Lambda(|\psi\rangle\langle\psi|)\cdot|\psi\rangle\langle\psi|\right]^2 \\
&= \sum_i \mathrm{tr}[K_i|\psi\rangle\langle\psi|]\mathrm{tr}[K_i^\dagger|\psi\rangle\langle\psi|]\sum_j \mathrm{tr}[K_j|\psi\rangle\langle\psi|]\mathrm{tr}[K_j^\dagger|\psi\rangle\langle\psi|] \\
&= \sum_{i,j} \mathrm{tr}\left[\left(K_i\otimes K_i^\dagger\otimes K_j\otimes K_j^\dagger\right)\cdot|\psi\rangle\langle\psi|^{\otimes 4}\right],
\end{aligned}$$

and then use Eq. (C.18) to simplify the average, $\overline{\mathcal{F}^2}$, as

$$\begin{aligned}
\overline{\mathcal{F}^2} &= \sum_{i,j} \mathrm{tr}\left[\left(K_i\otimes K_i^\dagger\otimes K_j\otimes K_j^\dagger\right)\overline{|\psi\rangle\langle\psi|^{\otimes 4}}\right] \\
&= \sum_{i,j} \mathrm{tr}\left[\left(K_i\otimes K_i^\dagger\otimes K_j\otimes K_j^\dagger\right)\frac{\pi_{\mathrm{sym}}(4,d)}{\mathrm{tr}\left[\pi_{\mathrm{sym}}(4,d)\right]}\right].
\end{aligned}$$

Finally, we write $\pi_{\mathrm{sym}}(4,d)$ as a sum of permutation operators

$$\overline{\mathcal{F}^2} = \frac{\sum_{i,j}\sum_{\sigma\in S_4}\mathrm{tr}\left[\left(K_i\otimes K_i^\dagger\otimes K_j\otimes K_j^\dagger\right)\mathcal{P}_\sigma\right]}{d(d+1)(d+2)(d+3)}, \tag{3.7}$$

invoke Eq. (C.20) to evaluate the traces, and use Eq. (C.19) to evaluate the normalization:

$$\overline{\mathcal{F}^2} = \frac{\sum_{i,j}\left(\mathrm{tr}\left[K_i\right]\mathrm{tr}\left[K_i^\dagger\right]\mathrm{tr}\left[K_j\right]\mathrm{tr}\left[K_j^\dagger\right] + \mathrm{tr}\left[K_iK_j^\dagger\right]\mathrm{tr}\left[K_i^\dagger\right]\mathrm{tr}\left[K_j\right] + ...\right)}{d(d+1)(d+2)(d+3)}. \tag{3.8}$$

There are 24 products of traces in the sum, each corresponding to one of the 4! permutations of 4 objects, so the ellipsis in the last equation represents 22 more terms.

It is more productive to use the basis $\{P_i\}$ since it has properties that allow for simpler calculation of traces. We write $\mathcal{F}$ using the expansion of the Kraus operators in terms of the χ matrix and the same calculation then yields

$$\overline{\mathcal{F}^2} = \frac{\sum_{l,m,n,r} \chi_{l,m}\chi_{n,r} \left(\mathrm{tr}\left[P_l\right]\mathrm{tr}\left[P_m\right]\mathrm{tr}\left[P_n\right]\mathrm{tr}\left[P_r\right] + \mathrm{tr}\left[P_l P_r\right]\mathrm{tr}\left[P_m\right]\mathrm{tr}\left[P_n\right] + ...\right)}{d(d+1)(d+2)(d+3)}. \tag{3.9}$$

By writing out all 24 terms in the summation (excluded here because of extreme tediousness), we can use the assumed properties of the basis $\{P_i\}$ to simplify this expression to

$$\overline{\mathcal{F}^2} = \frac{\left(\begin{array}{l} d^4\mathrm{tr}[\chi\chi_0]^2 \\ + d^3\mathrm{tr}\left[\chi_0\left(2\chi^2 + \chi\chi^T + \chi^T\chi + 2\chi\right)\right] \\ + d^2\left(4\mathrm{tr}[\chi\chi_0] + \mathrm{tr}\left[\chi\chi^T\right] + \mathrm{tr}\left[\chi^2\right] + 1\right) \\ + d\left(2\sum_l \mathrm{tr}\left[\left(\chi_{l,0} + \chi_{0,l}\right)P_l\Lambda\left(\mathbb{1}\right)\right] + 3\right) \\ + 2\mathrm{tr}\left[\sum_{l,m}\chi_{l,m}P_l\Lambda(P_m)\right] + \mathrm{tr}\left[\left(\Lambda\left(\mathbb{1}\right)\right)^2\right] \end{array}\right)}{d(d+1)(d+2)(d+3)} \tag{3.10}$$

where we have grouped terms in powers of d. All but three of the terms in Eq. (3.10) are expressed solely in terms of Λ's χ-matrix. The exceptions are:

- $\mathrm{tr}\left[\left(\Lambda\left(\mathbb{1}\right)\right)^2\right]$ which comes from terms of the form

$$\sum_{lmnr} \chi_{l,m}\chi_{n,r}\mathrm{tr}[P_l P_m P_n P_r] \tag{3.11}$$

  that are produced by 4-cycle permutations like $\sigma = (1234)$.

- $2d\mathrm{tr}\left[\sum_l \left(\chi_{l,0} + \chi_{0,l}\right)P_l\Lambda\left(\mathbb{1}\right)\right]$ which comes from terms of the form

$$\sum_{lmnr} \chi_{l,m}\chi_{n,r}\mathrm{tr}[P_l P_n P_r]\mathrm{tr}[P_m] \tag{3.12}$$

  that are produced by 3-cycle permutations like $\sigma = (123)(4)$.

- $2\mathrm{tr}\left[\sum_{l,m}\chi_{l,m}P_l\Lambda(P_m)\right]$ which comes from terms of the form

$$\sum_{lmnr} \chi_{l,m}\chi_{n,r}\mathrm{tr}[P_l P_n P_m P_r] \tag{3.13}$$

that are produced by 4-cycle permutations like $\sigma = (1324)$.

Our immediate goal is to use χ_0, the partial transpose, and the partial trace (see Sec. A.2) to rewrite these quantities in terms of the Jamiolkowski representation χ.

For the first term, it's straightforward to verify that

$$\Lambda\left(\frac{\mathbb{1}}{d}\right) = \mathrm{tr}_2\chi,$$

so

$$\mathrm{tr}\left[\left(\Lambda\left(\mathbb{1}\right)\right)^2\right] = d^2\mathrm{tr}\left[\left(\mathrm{tr}_2\chi\right)^2\right]. \tag{3.14}$$

We can rewrite the second term using the non-Hermitian operator

$$\begin{aligned}
\chi\chi_0 &= \sum_{l,m} \chi_{l,m}\left(P_l \otimes \mathbb{1}\right)\chi_0\left(P_m \otimes \mathbb{1}\right)\chi_0 \\
&= \sum_{l} \chi_{l,0}\left(P_l \otimes \mathbb{1}\right)\chi_0 \\
&= \frac{1}{d}\sum_{l,i,j} \chi_{l,0} P_l|i\rangle\langle j| \otimes |i\rangle\langle j|
\end{aligned}$$

and its adjoint $\chi_0\chi$. The second equality follows from the fact that

$$\begin{aligned}
\chi_0 P_m \otimes \mathbb{1}\chi_0 &= \frac{1}{d}\sum_{a,b,c,d} |a\rangle\langle b| \otimes |a\rangle\langle b|P_m \otimes \mathbb{1}|c\rangle\langle d| \otimes |c\rangle\langle d| \\
&= \frac{1}{d^2}\sum_{a,b,d} \langle b|P_m|b\rangle|a\rangle\langle d \otimes |a\rangle\langle d| \\
&= \delta_{m,0}\chi_0.
\end{aligned} \tag{3.15}$$

Partial tracing over the second (ancillary) system yields

$$\begin{aligned}\mathrm{tr}_2(\chi\chi_0) &= \frac{1}{d}\sum_l \chi_{l,0}P_l,\\ \mathrm{tr}_2(\chi_0\chi) &= \frac{1}{d}\sum_l \chi_{0,l}P_l,\end{aligned}$$

which provides the following expression for the second term:

$$\mathrm{tr}\left[\left(\sum_l \left(\chi_{l,0}+\chi_{0,l}\right)P_l\right)\Lambda\left(\mathbb{1}\right)\right] = d^2\mathrm{tr}\left[\mathrm{tr}_2\left(\chi\chi_0+\chi_0\chi\right)\mathrm{tr}_2\chi\right]. \tag{3.16}$$

To rewrite the third exceptional term, we apply a few more tricks. First, we observe that for any bipartite operator $A\otimes B$,

$$\mathrm{tr}\left[\chi^{T_2}(A\otimes B)\right] = \frac{1}{d}\mathrm{tr}\left[A\Lambda(B)\right]$$

which can be shown from the definition of χ through the following chain of equalitites,

$$\begin{aligned}\mathrm{tr}\left[\chi^{T_2}(A\otimes B)\right] &= \frac{1}{d}\mathrm{tr}\left(\sum_{c,d}\mathbb{1}\otimes|c\rangle\langle d|\left[\sum_{l,m}\Lambda(|l\rangle\langle m|)\otimes|l\rangle\langle m|\right]\otimes|c\rangle\langle d|(A\otimes B)\right)\\ &= \frac{1}{d}\sum_{l,m}\sum_k \mathrm{tr}\left(|l\rangle\langle m|A_k^\dagger AA_k\right)\mathrm{tr}\left(|m\rangle\langle l|B\right)\\ &= \frac{1}{d}\sum_{m,k}\mathrm{tr}\left(AA_kB|m\rangle\langle m|A_k^\dagger\right)\\ &= \frac{1}{d}\mathrm{tr}\left(A\Lambda(B)\right).\end{aligned} \tag{3.17}$$

Next, we note that since $\chi_0 = \frac{1}{d}\sum_{l,m}|l\rangle\langle m|\otimes|l\rangle\langle m|$, its partial transpose (over either subsystem) is

$$\chi_0^{T_1} = \chi_0^{T_2} = \frac{1}{d}\sum_{l,m}|l\rangle\langle m|\otimes|m\rangle\langle l|.$$

This bipartite operator is proportional to the unitary SWAP gate (which we denote S), which maps $|l\rangle\otimes|m\rangle\rightarrow|m\rangle\otimes|l\rangle$. Now, consider the operator $S(S\chi)^{T_1}$, which can be

written out in a simple form by first noting,

$$\begin{aligned}
S(S\chi)^{T_1} &= \frac{1}{d}S\left(S\sum_{ijlm}\chi_{lm}P_l|i\rangle\langle j|P_m\otimes|i\rangle\langle j|\right)^{T_1}\\
&= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}\left(\sum_{n,r}|n\rangle\langle r|\otimes|r\rangle\langle n|\right)P_l|i\rangle\langle j|P_m\otimes|i\rangle\langle j|\right)^{T_1}\\
&= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}\left(\sum_{r}|i\rangle\langle r|P_l|i\rangle\langle j|P_m\otimes|r\rangle\langle j\right)\right)^{T_1}\\
&= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}\left(\sum_{r}|i\rangle\langle j|P_m\otimes\langle r|P_l|i\rangle|r\rangle\langle j|\right)\right)^{T_1}\\
&= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}\left(|i\rangle\langle j|P_m\otimes\sum_{r}|r\rangle\langle r|P_l|i\rangle\langle j|\right)\right)^{T_1}\\
&= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}\left(|i\rangle\langle j|P_m\otimes P_l|i\rangle\langle j|\right)\right)^{T_1}.
\end{aligned}$$

Hence

$$\begin{aligned}
S(S\chi)^{T_1} &= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}\left(|i\rangle\langle j|P_m\otimes P_l|i\rangle\langle j|\right)\right)^{T_1}\\
&= \frac{1}{d}S\left(\sum_{ijlm}\chi_{lm}P_m^T|j\rangle\langle i|\otimes P_l|i\rangle\langle j|\right)\\
&= \frac{1}{d}\sum_{ijlm}\chi_{lm}\left(\sum_{n,r}|n\rangle\langle n|P_l|i\rangle\langle i|\otimes|r\rangle\langle r|P_m^T|j\rangle\langle j|\right)\\
&= \frac{1}{d}\sum_{ijlm}\chi_{lm}P_l|i\rangle\langle i|\otimes P_m^T|j\rangle\langle j|\\
&= \frac{1}{d}\sum_{lm}\chi_{lm}P_l\otimes P_m^T.
\end{aligned}$$

Together, these two observations imply that

$$\mathrm{tr}\left[\chi^{T_2}\left(S(S\chi)^{T_1}\right)^{T_2}\right] = \frac{1}{d^2}\sum_{lm}\chi_{lm}\mathrm{tr}[P_l\Lambda(P_m)], \tag{3.18}$$

but $\mathrm{tr}\left[X^{T_2}Y^{T_2}\right] = \mathrm{tr}\left[XY\right]$ (just as for the full transpose), so the two partial transposes cancel. Substituting in $S = d\chi_0^{T_1}$, we get the following expression for the third term:

$$\begin{aligned}\sum_{l,m}\chi_{l,m}\mathrm{tr}\left[P_l\Lambda(P_m)\right] &= d^4\mathrm{tr}\left[\chi\chi_0^{T_1}(\chi_0^{T_1}\chi)^{T_1}\right] \\ &= d^4\mathrm{tr}\left[\left(\chi_0^{T_1}\chi\right)^\dagger\left(\chi_0^{T_1}\chi\right)^{T_1}\right] \\ &= d^4\mathrm{tr}\left[\left(\chi\chi_0^{T_2}\right)^\dagger\left(\chi\chi_0^{T_2}\right)^{T_2}\right].\end{aligned} \tag{3.19}$$

Hence if,

$$\begin{aligned}a_1 &= \mathrm{tr}\left(\chi\chi_0\right)^2 + 2\mathrm{tr}\left[\left(\chi\chi_0^{T_2}\right)^\dagger\left(\chi\chi_0^{T_2}\right)^{T_2}\right], \\ b_1 &= 2\mathrm{tr}\left(\chi^2\chi_0\right) + \mathrm{tr}\left(\chi\chi^T\chi_0\right) + \mathrm{tr}\left(\chi^T\chi\chi_0\right) + 2\mathrm{tr}\left(\chi\chi_0\right) + 2\mathrm{tr}\left[\mathrm{tr}_2\left(\chi\chi_0 + \chi_0\chi\right)\mathrm{tr}_2\left(\chi\right)\right], \\ c_1 &= 4\mathrm{tr}\left(\chi\chi_0\right) + \mathrm{tr}\left(\chi\chi^T\right) + \mathrm{tr}\left(\chi^2\right) + 1 + \mathrm{tr}\left[\left(\mathrm{tr}_2\chi\right)^2\right], \\ d_1 &= 3,\end{aligned}$$

then,

$$\overline{\mathcal{F}^2} = \frac{a_1d^4 + b_1d^3 + c_1d^2 + d_1d}{d^4 + 6d^3 + 11d^2 + 6d}. \tag{3.20}$$

From Eq. (3.6) we have,

$$\overline{\mathcal{F}}^2 = \frac{ad^2 + bd + 1}{d^2 + 2d + 1} \tag{3.21}$$

where,

$$\begin{aligned} a &= \mathrm{tr}\left(\chi\chi_0\right)^2, \\ b &= 2\mathrm{tr}\left(\chi\chi_0\right). \end{aligned}$$

Taken together, Eq.'s (3.20) and (3.21) give the following expression for $\mathrm{Var}\left(\mathcal{F}\right)$,

$$\mathrm{Var}\left(\mathcal{F}\right) = \frac{a_2 d^5 + b_2 d^4 + c_2 d^3 + d_2 d^2 + e_2 d + f_2}{(d+1)^3(d+2)(d+3)} \tag{3.22}$$

where,

$$\begin{aligned} a_2 &= a_1 - a, \\ b_2 &= b_1 + 2a_1 - b - 6a, \\ c_2 &= a_1 + 2b_1 + c_1 - 11a - 6b - 1, \\ d_2 &= b_1 + 2c_1 + d_1 - 6a - 11b - 6 \\ e_2 &= c_1 + 2d_1 - 11 - 6b \\ f_2 &= d_1 - 6. \end{aligned}$$

3.3 Higher Order Moments

We briefly discuss how to calculate both the higher order moments $\overline{\mathcal{F}^m}$ and central moments $\overline{\left(\mathcal{F} - \overline{\mathcal{F}}\right)^m}$ of the gate fidelity $\mathcal{F}$. We have already given a detailed analysis of the $m = 1$ and $m = 2$ cases. More precisely we have provided explicit expressions for $\overline{\mathcal{F}}$, $\overline{\mathcal{F}^2}$, and $\mathrm{Var}(\mathcal{F}) = \overline{\mathcal{F}^2} - \overline{\mathcal{F}}^2$ in terms of the χ-matrix (Jamiolkowski state) of a quantum operation (note that the first central moment is just $\overline{\mathcal{F}}$). The central moments contain valuable information about the distribution of the gate fidelity. The second central moment (variance) is a measure of the spread of the distribution, the third central moment measures the skewness, and so on. Since the m'th central moment is just $\overline{\left(\mathcal{F} - \overline{\mathcal{F}}\right)^m}$ and we have an expression for $\overline{\mathcal{F}}$, the expression for the m'th central moment is easily obtained if each $\overline{\mathcal{F}^k}$ is known for $k = 1, ..., m$.

For $m \in \mathbb{N}$, the m'th power of $\mathcal{F}$, $\mathcal{F}^m$, has action on pure state $|\psi\rangle\langle\psi|$,

$$\begin{aligned}
\mathcal{F}^m\left(|\psi\rangle\langle\psi|\right) &= \left(\mathrm{tr}\left[\Lambda\left(|\psi\rangle\langle\psi|\right)|\psi\rangle\langle\psi|\right]\right)^m \\
&= \sum_{i_1} \mathrm{tr}\left[\left(K_{i_1}\otimes K_{i_1}^\dagger\right)|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right] ... \sum_{i_m} \mathrm{tr}\left[\left(K_{i_m}\otimes K_{i_m}^\dagger\right)|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right] \\
&= \sum_{i_1,...,i_m} \mathrm{tr}\left[\left(K_{i_1}\otimes K_{i_1}^\dagger\otimes ...\otimes K_{i_m}\otimes K_{i_m}^\dagger\right)\left(|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right)^{\otimes m}\right].
\end{aligned}$$

In an analogous method to that used in calculating an expression for the variance we have $\overline{\mathcal{F}^m}$ is given by,

$$\frac{\sum_{i_1,...,i_m} \mathrm{tr}\left[\left(K_{i_1}\otimes K_{i_1}^\dagger\otimes ...\otimes K_{i_m}\otimes K_{i_m}^\dagger\right)\pi_{sym}\left(2m,d\right)\right]}{\mathrm{tr}\left[\pi_{sym}\left(2m,d\right)\right]},$$

and using the results regarding permutation operators and the symmetric subspace described in Sec. C.3 we obtain,

$$\overline{\mathcal{F}^m} = \frac{\sum_{i_1,...,i_m}\left\{\mathrm{tr}\left(K_{i_1}\right)\mathrm{tr}\left(K_{i_1}^\dagger\right)...\mathrm{tr}\left(K_{i_m}\right)\mathrm{tr}\left(K_{i_m}^\dagger\right)+...+\mathrm{tr}\left(K_{i_1}K_{i_1}^\dagger...K_{i_m}K_{i_m}^\dagger\right)\right\}}{(2m)!\binom{2m+d-1}{d-1}} \tag{3.23}$$

where again the $\{K_i\}$ are a set of Kraus operators for Λ. There are $(2m)!$ terms in the sum corresponding to the fact that there are $(2m)!$ elements in the symmetric group S_{2m}. We have also used the fact that,

$$\mathrm{tr}\left[\pi_{sym}\left(2m,d\right)\right] = \binom{2m+d-1}{d-1}.$$

Expanding the K_i in terms of the basis $\{P_i\}$ with the previously discussed properties gives,

$$\overline{\mathcal{F}^m} = \frac{\sum\limits_{i_{1_1},i_{1_2},..,i_{m_1},i_{m_2}} \prod\limits_{j=1}^{m} \chi_{i_{j_1},i_{j_2}} \left\{ \mathrm{tr}\left(P_{i_{1_1}}\right) \mathrm{tr}\left(P_{i_{1_2}}\right) ... \mathrm{tr}\left(P_{i_{m_1}}\right) \mathrm{tr}\left(P_{i_{m_2}}\right) + ... \right\}}{(2m)! \binom{2m+d-1}{d-1}} \tag{3.24}$$

which can be written in terms of χ,

$$\overline{\mathcal{F}^m} = \frac{\left(\mathrm{tr}\left(\chi\chi_0\right)^m d^{2m} + ...\right)}{(2m)! \binom{2m+d-1}{d-1}}.$$

3.4 The Single Qubit Case

In this section we analyze the behavior of $\mathrm{Var}\left(\mathcal{F}\right)$ in the case of a single qubit ($d = 2$). For a qubit system, one can obtain much simpler equations for $\mathrm{Var}\left(\mathcal{F}\right)$ than Eq. (3.22) by using a different method of calculation (moreover, the expression given by Eq. (3.22) does not appear to simplify to the one obtained here). The calculation involves starting from Eq. (3.9), grouping certain terms together, and considering various cases. The result of the calculation is that the second moment of $\mathcal{F}$ is given by,

$$\overline{\mathcal{F}^2} = \frac{\left(\begin{array}{l} -48\mathrm{tr}\left(\chi\chi_0\right)^2 + 64\mathrm{tr}\left(\chi\chi_0\right) \\ + 24\mathrm{tr}\left(\chi\chi^T\chi_0 + \chi^T\chi\chi_0\right) + 32\mathrm{tr}\left(\chi^2\chi_0\right) \\ + 4\mathrm{tr}\left(\chi\chi^T\right) + 12\mathrm{tr}\left(\chi^2\right) + 4\mathrm{tr}\left[\left(\mathrm{tr}_2\chi\right)^2\right] + 6 \end{array} \right)}{120}.$$

Using Eq. (3.21) we obtain the following particularly simple analogue of Eq. (3.22),

$$\begin{aligned} \mathrm{Var}\left(\mathcal{F}\right) &= -\frac{11}{180} + \frac{4}{45}\mathrm{tr}\left(\chi\chi_0\right) - \frac{38}{45}\mathrm{tr}\left(\chi\chi_0\right)^2 + \frac{4}{15}\mathrm{tr}\left(\chi^2\chi_0\right) + \frac{1}{10}\mathrm{tr}\left(\chi^2\right) \\ &\quad + \frac{1}{5}\mathrm{tr}\left(\chi\chi^T\chi_0 + \chi^T\chi\chi_0\right) + \frac{1}{30}\left(\mathrm{tr}\left(\chi\chi^T\right) + \mathrm{tr}\left[\left(\mathrm{tr}_2\chi\right)^2\right]\right). \end{aligned} \tag{3.25}$$

Note that if $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$ is a Pauli channel then χ is diagonal and the variance takes the form,

$$\mathrm{Var}\left(\mathcal{F}\right) = -\frac{2}{45} + \frac{4}{45}\mathrm{tr}\left(\chi\chi_0\right) - \frac{8}{45}\mathrm{tr}\left(\chi\chi_0\right)^2 + \frac{2}{15}\mathrm{tr}\left(\chi^2\right).$$

The calculation for the single qubit case proceeds as follows. First, note that since we already have a simple expression for $\overline{\mathcal{F}}$ given by Eq. (3.6) we only need to calculate $\overline{\mathcal{F}^2}$. We will use Eq. (3.9) which will allow us to group particular terms together to obtain a more simple expression.

To begin with the calculation of $\overline{\mathcal{F}^2}$, we recall some properties of χ. First, χ is positive semi-definite and has trace equal to 1. Second,

$$\sum_{l,m} \chi_{l,m} P_l P_m = \Lambda(\mathbb{1}) = d\Lambda\left(\frac{\mathbb{1}}{d}\right)$$

and third,

$$\sum_{l,m} \chi_{l,m} P_m P_l = \mathbb{1}$$

from trace preservation. The 24 terms in Eq. (3.9) are sorted into groups of 3 each of which is dealt with separately. Since we are working with a single qubit, $d = 2$ in all expressions below. Note that many of the expressions below only hold under the assumption that $d = 2$.

3.4.1 First Group of Terms

The first group consists of the following 10 terms:

$$\sum_{l,m,n,r} \chi_{l,m}\chi_{n,r}([P_l][P_m][P_n][P_r] + [P_lP_r][P_m][P_n] + [P_mP_r][P_l][P_n] + [P_l][P_m][P_nP_r] + [P_lP_n][P_m][P_r]$$
$$+[P_lP_n][P_mP_r] + [P_mP_n][P_l][P_r] + [P_mP_n][P_lP_r] + [P_lP_m][P_n][P_r] + [P_lP_m][P_nP_r])$$

where for ease of presentation we have used square brackets "[ ]" to represent the trace operation. Using the assumed properties of the $\{P_i\}$ basis this group can be written as,

$$16\left(\chi_{0,0}^2+\sum_l \chi_{0,l}\chi_{l,0}+\chi_{0,0}\right)+8\left(\sum_l\left(\chi_{0,l}^2+\chi_{l,0}^2\right)\right)+4\left(\sum_{l,m}\left(\chi_{l,m}^2+\chi_{l,m}\chi_{m,l}\right)+1\right). \tag{3.26}$$

3.4.2 Second Group of Terms

The second group consists of the following 8 terms which are grouped as 4 pairs,

$$\begin{aligned}\sum_{l,m,n,r}\chi_{l,m}\chi_{n,r}&(\mathrm{tr}(P_lP_rP_n)\mathrm{tr}(P_m)+\mathrm{tr}(P_lP_nP_r)\mathrm{tr}(P_m)\\&+\mathrm{tr}(P_mP_rP_n)\mathrm{tr}(P_l)+\mathrm{tr}(P_mP_nP_r)\mathrm{tr}(P_l)\\&+\mathrm{tr}(P_rP_mP_l)\mathrm{tr}(P_n)+\mathrm{tr}(P_rP_lP_m)\mathrm{tr}(P_n)\\&+\mathrm{tr}(P_nP_mP_l)\mathrm{tr}(P_r)+\mathrm{tr}(P_nP_lP_m)\mathrm{tr}(P_r)).\end{aligned}$$

The four sums (one for each pair) are calculated independently. For the first sum we deal with five cases:

Case 1: $n\neq r$, $n\neq 0$, and $r\neq 0$. This implies $P_nP_r=-P_rP_n$ and so the above is 0.

Case 2: $n=r$. We get $2\sum_{l,m,n}\chi_{l,m}\chi_{n,n}\mathrm{tr}\left(P_l\right)\mathrm{tr}\left(P_m\right)$ which equals $2\chi_{0,0}d^2$.

Case 3: $n=0$. We get, $2\sum_{l,m,r}\chi_{l,m}\chi_{0,r}\mathrm{tr}\left(P_lP_r\right)\mathrm{tr}\left(P_m\right)$ which is just $2\sum_l\chi_{l,0}\chi_{0,l}d^2$.

Case 4: $r=0$. Similarly to case 3 we get $2\sum_l\chi_{l,0}^2d^2$.

Case 5: $r=0$ and $n=0$. This case is required because we have over-counted for this case twice above. The result is $2\chi_{0,0}^2d^2$.

Hence the five cases give that the first sum is equal to,

$$2\chi_{0,0}d^2 + 2\sum_l \chi_{l,0}\chi_{0,l}d^2 + 2\sum_l \chi_{l,0}^2 d^2 - 4\chi_{0,0}^2 d^2.$$

The other three sums are calculated in a similar fashion and in total the second group of terms is equal to,

$$8\chi_{0,0}d^2 + 8\sum_l \chi_{l,0}\chi_{0,l}d^2 + 4\sum_l \chi_{0,l}^2 d^2 + 4\sum_l \chi_{l,0}^2 d^2 - 16\chi_{0,0}^2 d^2.$$

Substituting $d = 2$ and collecting terms for both the first and second group of terms gives,

$$-48\chi_{0,0}^2 + 48\chi_{0,0} + 32\sum_l \chi_{0,l}^2 + 16\sum_l \chi_{l,0}^2 + 16\sum_l \chi_{l,0}\chi_{0,l} + 4\sum_{l,m} \chi_{l,m}^2 + 4\sum_{l,m} \chi_{l,m}\chi_{m,l} + 4. \tag{3.27}$$

3.4.3 Third Group of Terms

Lastly we have the following 6 terms which are grouped into 3 pairs,

$$\begin{aligned}\sum_{l,m,n,r} \chi_{l,m}\chi_{n,r}(&\mathrm{tr}\,(P_l P_r P_n P_m) + \mathrm{tr}\,(P_l P_r P_m P_n) \\ &+\mathrm{tr}\,(P_l P_n P_m P_r) + \mathrm{tr}\,(P_l P_m P_n P_r) \\ &+\mathrm{tr}\,(P_l P_m P_r P_n) + \mathrm{tr}\,(P_l P_n P_r P_m)).\end{aligned} \tag{3.28}$$

The first pair is easy to calculate using the same cases as above for m and n. The result is,

$$4\sum_{l,m} \chi_{l,m}\chi_{m,l} + 8\chi_{0,0} - 8\sum_l \chi_{l,0}\chi_{0,l}.$$

The second pair requires a bit more effort and we go through the cases separately,

Case 1: $m \neq n$, $m \neq 0$ and $n \neq 0$. This case gives 0.

Case 2: $m = n$. In this case the pair becomes $4\sum_{l,m} \chi_{l,m}\chi_{m,l}$.

Case 3: $m = 0$. The pair becomes $2\sum_{l,n,r} \chi_{l,0}\chi_{n,r}\mathrm{tr}\,(P_l P_n P_r)$ and after a direct calculation we get,

$$4\Big[\chi_{0,0} + \chi_{1,0}(\chi_{0,1} + \chi_{1,0} + i\chi_{2,3} - i\chi_{3,2}) + \chi_{2,0}(\chi_{0,2} + \chi_{2,0} - i\chi_{1,3} + i\chi_{3,1}) \\ +\chi_{3,0}(\chi_{0,3} + \chi_{3,0} + i\chi_{1,2} - i\chi_{2,1})\Big].$$

Case 4: $n = 0$. Similar to case 3 we obtain,

$$4\Big[\chi_{0,0} + \chi_{0,1}(\chi_{0,1} + \chi_{1,0} + i\chi_{2,3} - i\chi_{3,2}) + \chi_{0,2}(\chi_{0,2} + \chi_{2,0} - i\chi_{1,3} + i\chi_{3,1} \\ +\chi_{0,3}(\chi_{0,3} + \chi_{3,0} + i\chi_{1,2} - i\chi_{2,1})\Big].$$

Case 5: $m = 0$ and $n = 0$. This case gives $4\sum_l \chi_{l,0}\chi_{0,l}$.

Combining the 5 cases gives,

$$4\sum_{l,m} \chi_{l,m}\chi_{m,l} + 8\chi_{0,0} - 8\sum_l \chi_{0,l}\chi_{l,0} \tag{3.29}$$

$$+4\left(\chi_{0,1} + \chi_{1,0}\right)\left(\chi_{0,1} + \chi_{1,0} + i\left(\chi_{2,3} - \chi_{3,2}\right)\right) \tag{3.30}$$

$$+4\left(\chi_{0,2} + \chi_{2,0}\right)\left(\chi_{0,2} + \chi_{2,0} + i\left(\chi_{3,1} - \chi_{1,3}\right)\right) \tag{3.31}$$

$$+4\left(\chi_{0,3} + \chi_{3,0}\right)\left(\chi_{0,3} + \chi_{3,0} + i\left(\chi_{1,2} - \chi_{2,1}\right)\right). \tag{3.32}$$

$$\tag{3.33}$$

The third pair can be expressed as,

$$\sum_{l,m,n,r} \chi_{l,m}\chi_{n,r}(\mathrm{tr}\left(P_l P_m P_r P_n\right) + \mathrm{tr}\left(P_l P_n P_r P_m\right)) = \mathrm{tr}\left(\Lambda^\dagger\left[\Lambda^\dagger\left(\mathbb{1}\right)\right]\right) + \mathrm{tr}\left(\Lambda\left[\Lambda\left(\mathbb{1}\right)\right]\right)$$
$$= 4$$

and so combining the three pairs gives,

$$8\sum_{l,m}\chi_{l,m}\chi_{m,l} + 16\chi_{0,0} - 16\sum_{l}\chi_{l,0}\chi_{0,l} + 4 + 4\left(\chi_{0,1}+\chi_{1,0}\right)\left(\chi_{0,1}+\chi_{1,0}+i\left(\chi_{2,3}-\chi_{3,2}\right)\right)$$
$$+4\left(\chi_{0,2}+\chi_{2,0}\right)\left(\chi_{0,2}+\chi_{2,0}+i\left(\chi_{3,1}-\chi_{1,3}\right)\right) + 4\left(\chi_{0,3}+\chi_{3,0}\right)\left(\chi_{0,3}+\chi_{3,0}+i\left(\chi_{1,2}-\chi_{2,1}\right)\right). \tag{3.34}$$

We can calculate another expression for the three pairs by noting that four of the terms in Eq. (3.28) can be written as,

$$\mathrm{tr}\left(\Lambda\left[\Lambda^\dagger\left(\mathbb{1}\right)\right]\right) + \mathrm{tr}\left(\Lambda^\dagger\left[\Lambda\left(\mathbb{1}\right)\right]\right) + \mathrm{tr}\left(\Lambda\left[\Lambda\left(\mathbb{1}\right)\right]\right) + \mathrm{tr}\left(\Lambda^\dagger\left[\Lambda^\dagger\left(\mathbb{1}\right)\right]\right)$$

which is just $3d + \mathrm{tr}\left(\Lambda^\dagger\left[\Lambda\left(\mathbb{1}\right)\right]\right)$, or even more simply, $6 + \mathrm{tr}\left(\left[\Lambda\left(\mathbb{1}\right)\right]^2\right)$. The remaining two terms

$$\sum_{l,m,n,r}\chi_{l,m}\chi_{n,r}\mathrm{tr}\left(P_l P_n P_m P_r\right)$$

and

$$\sum_{l,m,n,r}\chi_{l,m}\chi_{n,r}\mathrm{tr}\left(P_l P_r P_m P_n\right)$$

are complex conjugates of one another. From the calculation of the first pair given above (see Eq. (3.4.3)),

$$\sum_{l,m,n,r}\chi_{l,m}\chi_{n,r}(\mathrm{tr}\left(P_l P_r P_n P_m\right) + \mathrm{tr}\left(P_l P_r P_m P_n\right)) = 4\sum_{l,m}\chi_{l,m}\chi_{m,l} + 8\chi_{0,0} - 8\sum_{l}\chi_{l,0}\chi_{0,l},$$

and since $\sum_{l,m,n,r} \chi_{l,m}\chi_{n,r}\mathrm{tr}\left(P_l P_r P_n P_m\right) = 2$,

$$\sum_{l,m,n,r} \chi_{l,m}\chi_{n,r}\mathrm{tr}\left(P_l P_r P_m P_n\right) = 4\sum_{l,m} \chi_{l,m}\chi_{m,l} + 8\chi_{0,0} - 8\sum_{l} \chi_{l,0}\chi_{0,l} - 2.$$

Therefore the three pairs can also be written as

$$2 + \mathrm{tr}\left(\left[\Lambda\left(\mathbb{1}\right)\right]^2\right) + 8\sum_{l,m} \chi_{l,m}\chi_{m,l} + 16\chi_{0,0} - 16\sum_{l} \chi_{l,0}\chi_{0,l}. \tag{3.35}$$

This gives an expression for $\mathrm{tr}\left(\left[\Lambda\left(\mathbb{1}\right)\right]^2\right)$ in terms of χ-matrix elements. Indeed by Eq.'s (3.14), (3.4.3) and (3.35),

$$\begin{aligned}
\mathrm{tr}\left(\left[\Lambda\left(\mathbb{1}\right)\right]^2\right) &= 4\mathrm{tr}\left(\left[\mathrm{tr}_2\chi\right]^2\right) \\
&= 2 + 4\left(\chi_{0,1} + \chi_{1,0}\right)\left(\chi_{0,1} + \chi_{1,0} + i\left(\chi_{2,3} - \chi_{3,2}\right)\right) \\
&\quad +4\left(\chi_{0,2} + \chi_{2,0}\right)\left(\chi_{0,2} + \chi_{2,0} + i\left(\chi_{3,1} - \chi_{1,3}\right)\right) \\
&\quad +4\left(\chi_{0,3} + \chi_{3,0}\right)\left(\chi_{0,3} + \chi_{3,0} + i\left(\chi_{1,2} - \chi_{2,1}\right)\right).
\end{aligned}$$

Combining all 24 terms given in Eq.'s (3.26), (3.27) and (3.35), using Eq. (3.14), simplifying sums, and noting $\mathrm{tr}\left(\pi_{\mathrm{sym}}(4,d)\right) = \frac{120}{24}$, we get,

$$\begin{aligned}
\overline{\mathcal{F}^2} &= \frac{-48\chi_{0,0}^2 + 64\chi_{0,0} + 24\left(\chi\chi^T + \chi^T\chi\right)_{0,0}}{120} \\
&+ \frac{32\left(\chi^2\right)_{0,0} + 4\mathrm{tr}\left(\chi\chi^T\right) + 12\mathrm{tr}\left(\chi^2\right) + 6 + 4\mathrm{tr}\left(\left(\mathrm{tr}_2\chi\right)^2\right)}{120}.
\end{aligned} \tag{3.36}$$

Using the definition of χ_0, and using the expression for the average fidelity given by Eq. (3.6), we have that the variance of the gate fidelity for a single qubit is given by Eq. (3.25).

3.5 Upper Bounds on the Variance

It is relatively straightforward to obtain a generic upper-bound on $\mathrm{Var}(\mathcal{F})$ that holds for any d and allows us to deduce the behavior of $\mathrm{Var}(\mathcal{F})$ in large dimensions. The idea is to use a suitable expression for the variance and bound the coefficients of the powers of d. The result is that,

$$\mathrm{Var}(\mathcal{F}) \leq \frac{4d^3 + 4d^{\frac{5}{2}} + 9d^2 + 4d^{\frac{3}{2}} + 5d}{(d+1)^2 (d+2)(d+3)}. \tag{3.37}$$

As a simple corollary, comparing powers of d in the numerator and denominator of Eq. (3.37), we see that for large d,

$$\mathrm{Var}(\mathcal{F}) \;\sim\; O\left(\frac{1}{d}\right). \tag{3.38}$$

We again emphasize that Eq. (3.37) is completely general: for *any* quantum operation $\mathcal{E}$ and *any* unitary operation $\mathcal{U}$ acting on $L\left(\mathbb{C}^d\right)$, the gate fidelity between $\mathcal{E}$ and $\mathcal{U}$ has variance that satisfies Eq. (3.37).

To deduce the asymptotic behavior of $\mathrm{Var}(\mathcal{F})$ given by Eq. (3.37) we use the expression for $\overline{\mathcal{F}^2}$ given in Eq. (3.10). From this equation one can obtain the following expresssion for $\mathrm{Var}(\mathcal{F})$,

$$\mathrm{Var}(\mathcal{F}) = \frac{rd^4 + sd^3 + ud^2 + vd + w}{d(d^2 + 2d + 1)(d^2 + 5d + 1)}. \tag{3.39}$$

where,

$$
\begin{aligned}
r &= -4\chi_{0,0}^2 + \left(\chi\chi^T\right)_{0,0} + \left(\chi^T\chi\right)_{0,0} + 2\left(\chi^2\right)_{0,0}, \\
s &= -6\chi_{0,0}^2 + \left(\chi\chi^T\right)_{0,0} + \left(\chi^T\chi\right)_{0,0} + \mathrm{tr}\left(\chi\chi^T\right) - 4\chi_{0,0} + \mathrm{tr}\left(\chi^2\right) + 2\left(\chi^2\right)_{0,0}, \\
u &= -8\chi_{0,0} + \mathrm{tr}\left(\chi\chi^T\right) + \mathrm{tr}\left(\chi^2\right) + 2\mathrm{tr}\left(\sum_l \left(\chi_{l,0} + \chi_{0,l}\right) P_l \Lambda\left(\mathbb{1}\right)\right) - 1, \\
v &= 2\mathrm{tr}\left(\sum_{l,m} \chi_{l,m} P_l \Lambda(P_m)\right) + 2\mathrm{tr}\left(\sum_l \left(\chi_{l,0} + \chi_{0,l}\right) P_l \Lambda\left(\mathbb{1}\right)\right) + \mathrm{tr}\left(\left(\Lambda\left(\mathbb{1}\right)\right)^2\right) - 3, \\
w &= 2\mathrm{tr}\left(\sum_{l,m} \chi_{l,m} P_l \Lambda(P_m)\right) + \mathrm{tr}\left(\left(\Lambda\left(\mathbb{1}\right)\right)^2\right).
\end{aligned}
$$

The denominator of (3.39) is a quintic polynomial in d. The numerator contains powers of d up to and including d^4, however the coefficients depend on χ. We would like to bound these coefficients in terms of d.

First, since χ is a trace-1 positive semi-definite matrix, we obtain the bounds

$$0 \leq \chi_{0,0}^2 \leq \chi_{0,0} \leq 1 \tag{3.40}$$

and

$$0 \leq \mathrm{tr}\left(\chi^2\right) \leq \mathrm{tr}\left(\chi\right) = 1. \tag{3.41}$$

Next, for a linear operator A, the Frobenius (Hilbert-Schmidt) norm of A, denoted by $\| \ \|_F$, is given by $\|A\|_F = \sqrt{\mathrm{tr}\left(A^\dagger A\right)}$. Using the Cauchy-Schwarz inequality we obtain,

$$\left|\mathrm{tr}\left(\chi\chi^T\right)\right| \leq \|\chi\|_F \left\|\chi^T\right\|_F.$$

Since χ and χ^T have the same singular values,

$$\|\chi\|_F = \|\chi^T\|_F. \tag{3.42}$$

Therefore $\|\chi\|_F \leq 1 \Rightarrow \left|\mathrm{tr}\left(\chi\chi^T\right)\right| \leq 1$. This also implies

$$\left|\left(\chi\chi^T\right)_{0,0}\right| \leq 1 \text{ and } \left|\left(\chi^T\chi\right)_{0,0}\right| \leq 1. \tag{3.43}$$

To deal with $\Lambda(\mathbb{1})$, we note that it has trace d and is positive semi-definite. Hence

$$0 \leq \operatorname{tr}\left(\left(\Lambda(\mathbb{1})\right)^2\right) \leq d^2. \tag{3.44}$$

The only two coefficients that remain to be bounded are

$$\operatorname{tr}\left(\sum_l \left(\chi_{l,0}+\chi_{0,l}\right) P_l \Lambda(\mathbb{1})\right) \text{ and } \operatorname{tr}\left(\sum_{l,m} \chi_{l,m} P_l \Lambda(P_m)\right). \tag{3.45}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\operatorname{tr}\left(\sum_l \chi_{l,0} P_l \Lambda(\mathbb{1})\right)\right| &\leq \left\|\sum_l \chi_{l,0} P_l\right\|_F \|\Lambda(\mathbb{1})\|_F \\ &\leq d\left\|\sum_l \chi_{l,0} P_l\right\|_F.\end{aligned}$$

Since,

$$\begin{aligned}\left\|\sum_l \chi_{l,0} P_l\right\|_F &= \sqrt{\operatorname{tr}\left(\left(\sum_l \chi_{l,0} P_l\right)^\dagger \left(\sum_m \chi_{m,0} P_m\right)\right)} \\ &= \sqrt{\operatorname{tr}\left(\left(\sum_l \chi_{0,l} P_l\right)\left(\sum_m \chi_{m,0} P_m\right)\right)} \\ &= \sqrt{d}\sqrt{\left(\chi^2\right)_{0,0}} \\ &\leq \sqrt{d}\end{aligned}$$

we get $\left|\operatorname{tr}\left(\sum_l \left(\chi_{l,0}+\chi_{0,l}\right) P_l \Lambda(\mathbb{1})\right)\right| \leq 2d^{\frac{3}{2}}$.

Finally, we need to bound $\mathrm{tr}\left(\sum_{l,m}\chi_{l,m}P_l\Lambda(P_m)\right)$. Using Eq. (3.18) we have

$$\mathrm{tr}\left(\sum_{l,m}\chi_{l,m}P_l\Lambda(P_m)\right) = d^2\mathrm{tr}\left(S\left(S\chi\right)^{T_1}\chi\right) \tag{3.46}$$

where we recall S is the unitary Kraus operator for the SWAP gate. Again, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|\mathrm{tr}\left(\chi S\left(S\chi\right)^{T_1}\right)\right| &\leq \left\|\chi S\right\|_F \left\|\left(S\chi\right)^{T_1}\right\|_F \\
&= \sqrt{\mathrm{tr}\left(\left(\chi S\right)\left(\chi S\right)^\dagger\right)}\sqrt{\mathrm{tr}\left(\left(\left(S\chi\right)^{T_1}\right)^\dagger\left(S\chi\right)^{T_1}\right)} \\
&\leq \sqrt{\mathrm{tr}\left(\left(\left(S\chi\right)^{T_1}\right)^\dagger\left(S\chi\right)^{T_1}\right)}
\end{aligned}$$

where the last inequality holds because,

$$\begin{aligned}
\sqrt{\mathrm{tr}\left(\left(\chi S\right)\left(\chi S\right)^\dagger\right)} &= \sqrt{\mathrm{tr}\left(\chi^2\right)} \\
&= \left\|\chi\right\|_F \\
&\leq 1.
\end{aligned} \tag{3.47}$$

Now for any A, $B \in L\left(\mathcal{H}\otimes\mathcal{H}\right)$,

$$\left(A^\dagger\right)^{T_1} = \left(A^{T_1}\right)^\dagger \text{ and } \mathrm{tr}\left(\left(AB\right)^{T_1}\right) = \mathrm{tr}\left(B^{T_1}A^{T_1}\right). \tag{3.48}$$

Therefore,

$$\begin{aligned}
\operatorname{tr}\left(\left((S\chi)^{T_1}\right)^{\dagger}(S\chi)^{T_1}\right) &= \operatorname{tr}\left(\left((S\chi)^{\dagger}\right)^{T_1}(S\chi)^{T_1}\right) \\
&= \operatorname{tr}\left((S\chi)^{T_1}\left((S\chi)^{\dagger}\right)^{T_1}\right) \\
&= \operatorname{tr}\left(\left((S\chi)^{\dagger}(S\chi)\right)^{T_1}\right) \\
&= \operatorname{tr}\left(\chi^2\right) \leq 1,
\end{aligned}$$

which implies $\left|\operatorname{tr}\left(\sum_{l,m}\chi_{l,m}P_l\Lambda(P_m)\right)\right| \leq d^2$.

Combining all of these results and ignoring negative terms in (3.39) gives,

$$\begin{aligned}
\operatorname{Var}(\mathcal{F}) &\leq \frac{|r|d^4+|s|d^3+|u|d^2+|v|d+|w|}{d\left(d^2+2d+1\right)\left(d^2+5d+6\right)} \\
&\leq \frac{4d^3+4d^{\frac{5}{2}}+9d^2+4d^{\frac{3}{2}}+5d}{(d+1)^2(d+2)(d+3)} \\
&\sim O\left(\frac{1}{d}\right). \qquad (3.49)
\end{aligned}$$

3.6 Statistical Properties and Asymptotic Behavior of the Gate Fidelity

The aim of this section is to deduce various statistical properties of the gate fidelity, many of which are asymptotic. This is done by viewing the gate fidelity as a random variable on $\mathbb{CP}^{d-1}$, where we assume $\mathbb{CP}^{d-1}$ is equipped with the Fubini-Study measure μ_{FS} [12]. Again by equation (A.75) there is no loss in generality in restricting attention to gate fidelities of the form $\mathcal{F}_{\Lambda,\mathcal{I}}$ where Λ is some quantum operation. We abandon the compact notation of "$\mathcal{F}$" from the previous section.

The variance of $\mathcal{F}_{\Lambda,\mathcal{I}}$, which for simplicity we denote by $\sigma^2(\Lambda)$, is given by

$$\sigma^2(\Lambda) = \mathbb{E}_{\mu_{FS}}\left[\left(\mathcal{F}_{\Lambda,\mathcal{I}}-\overline{\mathcal{F}_{\Lambda,\mathcal{I}}}\right)^2\right] = \overline{\mathcal{F}_{\Lambda,\mathcal{I}}^2}-\overline{\mathcal{F}_{\Lambda,\mathcal{I}}}^2.$$

If Λ is depolarizing then $\sigma^2(\Lambda) = 0$ and from Sec. 3.1, for d $\geq$ 4, a non-depolarizing quantum channel $\mathcal{R}$ was constructed which satisfies $\mathcal{F}_{\mathcal{R},\mathcal{I}} = \mathcal{F}_{\Lambda,\mathcal{I}}$. Hence there exist non-depolarizing quantum channels $\mathcal{R}$ with $\sigma^2(\mathcal{R}) = 0$. Therefore it is *not* true that Λ is a depolarizing channel if and only if the variance of $\mathcal{F}_{\Lambda,\mathcal{I}}$ is 0.

From Eq. (3.37), $\sigma^2(\Lambda) \to 0$ as $\frac{1}{d}$ when $d \to \infty$. Moreover this holds for *any* quantum channel Λ. Therefore for large d and any channel Λ, $\mathcal{F}_{\Lambda,\mathcal{I}}$ must be "close" to $\mathcal{F}_{\Lambda_{\text{dep}},\mathcal{I}}$ as random variables. This idea will be made precise in Sec. 3.6.4 using both a natural metric on ξ and bounds obtained in Sec. D.2.

3.6.1 Concentration of Measure for the Gate Fidelity

In this subsection, Levy's lemma (discussed in Sec. D.3) is used to make precise the idea that $\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi\rangle)$ is close to $\overline{\mathcal{F}_{\Lambda,\mathcal{I}}}$ when $|\phi\rangle$ is chosen uniformly at random according to μ_{FS}. The key is to show that $\mathcal{F}_{\Lambda,\mathcal{I}}$ satisfies a Lipschitz condition which is *independent* of the dimension d of the system.

Theorem 3. *The function* $\mathcal{F}_{\Lambda,\mathcal{I}} : (\mathbb{CP}^{d-1}, \|\ \|_2) \to [0,1]$ *is* $3\sqrt{2}$*-Lipschitz.*

Proof. The goal is to show that $\forall |\phi_1\rangle, |\phi_2\rangle \in \mathbb{CP}^{d-1}$,

$$|\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_1\rangle) - \mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_2\rangle)| \leq 3\sqrt{2}\||\phi_1\rangle - |\phi_2\rangle\|_2.$$

By the triangle inequality,

$$\begin{aligned} |\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_1\rangle) - \mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_2\rangle)| \quad \leq \quad & |\text{tr}\left(|\phi_1\rangle\langle\phi_1|\left(\Lambda\left(|\phi_1\rangle\langle\phi_1|\right) - \Lambda\left(|\phi_2\rangle\langle\phi_2|\right)\right)\right)| \\ & + |\text{tr}\left(\Lambda(|\phi_2\rangle\langle\phi_2|)\left(|\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|\right)\right)|. \end{aligned}$$

Let $\|\ \|_1$ and $\|\ \|_2$ be the Schatten 1 and 2-norms (ie. trace and Frobenius norms) on $L(\mathcal{H})$ respectively [142]. By the Cauchy-Schwarz inequality,

$$\begin{aligned} |\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_1\rangle) - \mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_2\rangle)| \quad \leq \quad & \||\phi_1\rangle\langle\phi_1|\|_2\|\Lambda(|\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|)\|_2 \\ & + \|\Lambda(|\phi_2\rangle\langle\phi_2|)\|_2\||\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|\|_2. \end{aligned}$$

For any linear operator $A \in L(\mathcal{H})$ [71],

$$\|A\|_2 \leq \|A\|_1 \leq \sqrt{\text{rank}(A)}\|A\|_2$$

which gives $\|\Lambda(|\phi_2\rangle\langle\phi_2|)\|_2 \leq \|\Lambda(|\phi_2\rangle\langle\phi_2|)\|_1 = 1$. As well for any pure state $|\psi\rangle\langle\psi|$, $\||\psi\rangle\langle\psi|\|_2 = 1$. Therefore,

$$|\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_1\rangle) - \mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_2\rangle)| \leq \|\Lambda(|\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|)\|_1 + \||\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|\|_2.$$

Using the fact that quantum operations can only decrease the $\|\ \|_1$ distance between quantum states [108] and also that the difference of two rank 1 projectors has rank at most 2,

$$|\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_1\rangle) - \mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_2\rangle)| \leq 3\||\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|\|_2.$$

Finally, the Frobenius norm needs to be related to the Euclidean distance between $|\phi_1\rangle$ and $|\phi_2\rangle$. Note that

$$\||\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|\|_2 = \sqrt{2}\sqrt{1 - |\langle\phi_1|\phi_2\rangle|^2}$$

and,

$$\begin{aligned} \||\phi_1\rangle - |\phi_2\rangle\|_2 &= \sqrt{(\langle\phi_1| - \langle\phi_2|)\,(|\phi_1\rangle - |\phi_2\rangle)} \\ &= \sqrt{2}\sqrt{1 - \text{Re}\,(\langle\phi_1|\phi_2\rangle)}. \end{aligned} \tag{3.50}$$

Hence,

$$\begin{aligned} \||\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2|\|_2 &\leq \sqrt{2}\sqrt{1 - \text{Re}\,(\langle\phi_1|\phi_2\rangle)}\sqrt{1 + \text{Re}\,(\langle\phi_1|\phi_2\rangle)} \\ &\leq \sqrt{2}\||\phi_1\rangle - |\phi_2\rangle\|_2. \end{aligned}$$

Therefore,

$$|\mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_1\rangle) - \mathcal{F}_{\Lambda,\mathcal{I}}(|\phi_2\rangle)| \leq 3\sqrt{2}\||\phi_1\rangle - |\phi_2\rangle\|_2, \tag{3.51}$$

and so $3\sqrt{2}$ is a Lipschitz constant for $\mathcal{F}_{\Lambda,\mathcal{I}} : (\mathbb{CP}^{d-1}, \| \|_2) \to \mathbb{R}$ which proves the theorem. □

For fixed d, the infimum over all K such that $\mathcal{F}_{\Lambda,\mathcal{I}}$ is K-Lipschitz is called the Lipschitz seminorm of $\mathcal{F}_{\Lambda,\mathcal{I}}$. If the Lipschitz seminorm of $\mathcal{F}_{\Lambda,\mathcal{I}}$ is denoted by η then an obvious corollary of the above theorem is that η is bounded above by $3\sqrt{2}$.

The metric space isomorphism between $(\mathbb{S}^{2d-1}, \| \|_2)$ and the set of unit vectors in $\mathbb{C}^d$ implies that the function $\mathcal{F}_{\Lambda,\mathcal{I}} : (\mathbb{S}^{2d-1}, \| \|_2) \to [0,1]$ is $3\sqrt{2}$-Lipschitz. As discussed in Sec. D.3 this implies that for $\epsilon > 0$,

$$\mathbb{P}_{\mu_{FS}}\left[\mathcal{F}_{\Lambda,\mathcal{I}} \in \left(\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon, \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} + \epsilon\right)\right] \geq 1 - 4e^{\frac{-d\epsilon^2}{81\pi^3 ln2}}. \tag{3.52}$$

Hence, if $\epsilon > 0$, and $|\phi\rangle$ is chosen randomly from the Fubini-Study measure, the probability that the fidelity between $\Lambda(|\phi\rangle\langle\phi|)$ and $|\phi\rangle\langle\phi|$ is not ϵ-close to the average is exponentially small in d, ie.

$$\mathrm{pr}\left[\mathrm{tr}\left(\Lambda\left(|\psi\rangle\langle\psi|\right)|\psi\rangle\langle\psi|\right) \in \left(\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon, \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} + \epsilon\right)\right] \geq 1 - 4e^{\frac{-d\epsilon^2}{81\pi^3 ln2}}. \tag{3.53}$$

3.6.2 Estimates and Bounds for the Average and Variance of the Gate Fidelity

The results of the previous section imply that the number of trials required to estimate the average gate fidelity between an unknown quantum operation Λ and $\mathcal{I}$ decreases significantly as d grows large. Unfortunately generating Haar-random pure states is an inefficient task. It would therefore be useful to derive deviation inequalities similar to those given above for discrete sets of states with the counting measure. A natural set of states to analyze in this context are state k-designs [122], in particular approximate state 1 and 2-designs due to their ability to be efficiently generated [8].

A state k-design consists of states spread uniformly enough throughout $\mathbb{CP}^{d-1}$ so that the k'th central moment of the gate fidelity over the t-design is equal to the k'th central moment over $\mathbb{CP}^{d-1}$. An approximate state k-design is a finite set of states that approximates the k'th central moment over $\mathbb{CP}^{d-1}$ well. From Eq. (3.53) it would be interesting

to investigate in large dimensions whether choosing a state uniformly at random from an approximate k-design also provides a good estimate of the average fidelity with high probability.

As mentioned previously, an explicit upper bound on the variance of Λ, denoted $\sigma^2(\Lambda)$, is given by Eq. (3.37) which shows that $\sigma^2(\Lambda)$ scales as $O\left(\frac{1}{d}\right)$. One can also use the concentration results derived above to deduce both the asymptotic order of $O\left(\frac{1}{d}\right)$ for $\sigma^2(\Lambda)$ as well as an explicit upper bound that holds for every d. The method has the advantage of not requiring an exact expression for the variance and therefore is much simpler to obtain. The downside is that the upper bound is not as tight. For simplicity, $\sigma^2(\Lambda)$ will be denoted by σ^2 below.

The asymptotic order of σ^2 is obtained by using Eq. (3.52) and Chebyshev's inequality which states that for any $k > 0$,

$$\mathbb{P}_{\mu_{FS}}\left[\mathcal{F}_{\Lambda,\mathcal{I}} \in \left(\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - k\sigma, \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} + k\sigma\right)\right] \geq 1 - \frac{1}{k^2}.$$

From Eq. (3.52), any $\sigma > 0$ that satisfies the above equation for all d and $k > 0$ must scale as $O\left(\frac{1}{\sqrt{d}}\right)$. More precisely, the right-hand side of Eq. (3.52) is constant if and only if ϵ scales as $O\left(\frac{1}{\sqrt{d}}\right)$. This is equivalent to $k\sigma = \epsilon \sim O\left(\frac{1}{\sqrt{d}}\right)$ and so if k is constant, $\sigma = \epsilon \sim O\left(\frac{1}{\sqrt{d}}\right)$. Therefore the variance σ^2 scales as $O\left(\frac{1}{d}\right)$.

For the upper bound on σ^2, if $\epsilon > 0$ let A_ϵ denote the set $\mathcal{F}_{\Lambda,\mathcal{I}} \in \left(\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon, \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} + \epsilon\right)$. An upper bound can be found by noting that for any $\epsilon > 0$,

$$\begin{aligned}
\sigma^2 &= \mathbb{E}_{\mu_{FS}}\left[\left(\mathcal{F}_{\Lambda,\mathcal{I}} - \overline{\mathcal{F}_{\Lambda,\mathcal{I}}}\right)^2 \mathbb{1}_{A_\epsilon}\right] + \mathbb{E}_{\mu_{FS}}\left[\left(\mathcal{F}_{\Lambda,\mathcal{I}} - \overline{\mathcal{F}_{\Lambda,\mathcal{I}}}\right)^2 \mathbb{1}_{\mathbb{CP}^{d-1}/A_\epsilon}\right] \\
&\leq \epsilon^2 + 4e^{\frac{-d\epsilon^2}{81\pi^3 ln2}}\left(\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon\right)^2 \\
&\leq \epsilon^2 + 4e^{\frac{-d\epsilon^2}{81\pi^3 ln2}}\left(1 - \epsilon\right)^2 \\
&\leq \epsilon^2 + 4e^{\frac{-d\epsilon^2}{81\pi^3 ln2}}. \qquad (3.54)
\end{aligned}$$

where $\mathbb{1}_A$ is the indicator function on $\mathbb{CP}^{d-1}$ with support on A, and similarly for $\mathbb{1}_{\mathbb{CP}^{d-1}/A}$. Minimizing with respect to ϵ and defining $C = \frac{1}{81\pi^3 \ln(2)}$ gives,

$$\epsilon = \sqrt{\frac{\ln(Cd)}{Cd}}.$$

Hence

$$\sigma^2 \leq \frac{4 + \ln(Cd)}{Cd}$$

and so for n qubits,

$$\sigma^2 \leq \frac{4 + \ln(C) + \frac{n}{\ln(2)}}{C2^n}.$$

As an example, for a 50 qubit system the above gives $\sigma^2 \leq 1.1 \times 10^{-10}$. On the other hand Eq. (3.37) gives a tighter bound of 1.0×10^{-14}. Clearly for systems capable of performing large-scale quantum computations the variance of the gate fidelity will be extremely small. Next we look at a specific example in the hope of gaining intuition as to how these results hold for any quantum channel Λ.

3.6.3 Amplitude Damping Channels

Amplitude damping is a useful example for understanding many of the results obtained to this point. For a single qubit, a Kraus representation for the amplitude damping channel $\mathcal{E}_{AD}$ is given by,

$$\mathcal{E}_{AD}(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger$$

where

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix} \text{ and } E_1 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}.$$

A physical application of the amplitude damping channel is in the description of spontaneous emission of a two-level atom coupled to a vacuum [23]. The parameter p in this setting is equal to the probability that the excited state decays to the ground state and when $p = 1$ every state is mapped to the $|0\rangle$ state.

Let us generalize the $p = 1$ case to higher dimensions. Suppose d is arbitrary, $p = 1$, and we fix an output state $|\psi\rangle \in \mathbb{CP}^{d-1}$. For every state $\rho \in L(\mathcal{H})$ we define the action of $\mathcal{E}_{AD}$ by

$$\mathcal{E}_{AD}(\rho) = |\psi\rangle\langle\psi|.$$

Since $\mathcal{F}_{\mathcal{E}_{AD},\mathcal{I}}(|\psi\rangle) = 1$ and for any $|\phi\rangle$ orthogonal to $|\psi\rangle$, $\mathcal{F}_{\mathcal{E}_{AD},\mathcal{I}}(|\phi\rangle) = 0$ it might appear that $\mathcal{F}_{\mathcal{E}_{AD},\mathcal{I}} : \mathbb{CP}^{d-1} \to [0,1]$ has a non-negligible variance for any d. From section 3.6.2 though it must be the case that the variance of $\mathcal{F}_{\mathcal{E}_{AD},\mathcal{I}}$ converges to 0 as $O\left(\frac{1}{d}\right)$. A simple result on the oblateness of the real unit sphere in large dimensions, which is related to the concentration of measure effect, provides an elegant explanation for why $\mathcal{E}_{AD}$ appears depolarizing with respect to the gate fidelity as $d \to \infty$.

It is a well-known fact that $\mathbb{S}^{2d-1}$ becomes oblate (appears to "fatten") about any equator with respect to the Haar measure [90]. More precisely, if C is a spherical cap on $\mathbb{S}^{2d-1}$ whose base is ϵ units of distance away from the origin then

$$\mu(C) \leq e^{-d\epsilon^2}.$$

Therefore the measure of two opposing spherical caps is bounded above by $2e^{-d\epsilon^2}$ and so the ϵ-neighborhood of any equator has measure at least $1 - 2e^{-d\epsilon^2}$.

Now $|\psi\rangle \in \mathbb{CP}^{d-1}$ can be associated with a point r on $\mathbb{S}^{2d-1}$ which defines a preferred direction on $\mathbb{S}^{2d-1}$ (ie. the north pole). Hence the association of $|\psi\rangle$ with r defines a unique equator on $\mathbb{S}^{2d-1}$. Choosing a state uniformly at random from the Fubini-Study measure on $\mathbb{CP}^{d-1}$ is equivalent to choosing a unit vector uniformly at random from $\mathbb{S}^{2d-1}$. Therefore by the oblateness of the unit sphere in large dimensions, if d is large, a state chosen uniformly at random on $\mathbb{CP}^{d-1}$ will, with exponentially high probability, lie close to the equator defined by r. By symmetry, every point on this equator will have the same value for the fidelity with $|\psi\rangle$. Therefore the value of the fidelity between a randomly chosen state and $|\psi\rangle$ will, with exponentially high probability, lie inside a small interval centered around the average. As a result, $\mathcal{F}_{\mathcal{E}_{AD},\mathcal{I}} : \mathbb{CP}^{d-1} \to [0,1]$ must have a small variance for large d.

3.6.4 Convergence to Depolarization

This section will bring together many of the results from the previous sections as a single result: the asymptotic convergence to depolarization of quantum channels with respect

to the gate fidelity. The convergence is quantified in two ways, the first utilizing the L^2 metric on the set ξ of gate fidelity random variables and the second resembling the notion of convergence in probability.

If $\mathcal{G}$ and $\mathcal{K}$ are two quantum operations on $L(\mathcal{H})$ then the L^2 distance, denoted here by d_2, between $\mathcal{F}_{\mathcal{G},\mathcal{I}}$ and $\mathcal{F}_{\mathcal{K},\mathcal{I}}$ is,

$$d_2\left(\mathcal{F}_{\mathcal{G},\mathcal{I}}, \mathcal{F}_{\mathcal{K},\mathcal{I}}\right) = \left(\mathbb{E}_{\mu_{FS}}\left[\left(\mathcal{F}_{\mathcal{G},\mathcal{I}} - \mathcal{F}_{\mathcal{K},\mathcal{I}}\right)^2\right]\right)^{\frac{1}{2}}.$$

Suppose that $\mathcal{G}$ has average fidelity equal to b and that $\mathcal{K}$ is the depolarizing channel with (constant) gate fidelity equal to b. Denoting $\mathcal{K}$ by $\mathcal{G}_{\text{dep}}$,

$$d_2\left(\mathcal{F}_{\mathcal{G},\mathcal{I}}, \mathcal{F}_{\mathcal{G}_{\text{dep}},\mathcal{I}}\right) = \left(\mathbb{E}_{\mu_{FS}}\left[\left(\mathcal{F}_{\mathcal{G},\mathcal{I}} - b\right)^2\right]\right)^{\frac{1}{2}}$$

which is just the standard deviation of $\mathcal{F}_{\mathcal{G},\mathcal{I}}$. Therefore from Eq. (3.37), for every d,

$$d_2\left(\mathcal{F}_{\mathcal{G},\mathcal{I}}, \mathcal{F}_{\mathcal{G}_{\text{dep}},\mathcal{I}}\right) \leq \sqrt{\frac{4d^3 + 4d^{\frac{5}{2}} + 9d^2 + 4d^{\frac{3}{2}} + 5d}{(d+1)^2(d+2)(d+3)}} \tag{3.55}$$

and so $d_2\left(\mathcal{F}_{\mathcal{G},\mathcal{I}}, \mathcal{F}_{\mathcal{G}_{\text{dep}},\mathcal{I}}\right) \to 0$ as $O\left(\frac{1}{\sqrt{d}}\right)$.

The second method uses the concentration of measure results from Sec. D.2. It is straightforward to turn Eq. (3.52) into a statement regarding convergence to depolarization by noting that since $\mathcal{F}_{\mathcal{G}_{\text{dep}},\mathcal{I}}$ is constant and equal to b, for any $\epsilon > 0$,

$$\mathbb{P}\left[\left|\mathcal{F}_{\mathcal{G},\mathcal{I}} - \mathcal{F}_{\mathcal{G}_{\text{dep}},\mathcal{I}}\right| \leq \epsilon\right] \geq 1 - 4e^{\frac{-d\epsilon^2}{81\pi^3 \ln(2)}}.$$

Hence for $\epsilon > 0$ fixed,

$$\lim_{d\to\infty} \mathbb{P}\left[\left|\mathcal{F}_{\mathcal{G},\mathcal{I}} - \mathcal{F}_{\mathcal{G}_{\text{dep}},\mathcal{I}}\right| \leq \epsilon\right] = 1.$$

3.6.5 Estimating the Minimum Gate Fidelity

In this section methods are discussed for estimating the minimum of the gate fidelity. The first method uses the Lipschitz constant given by Eq. (3.51) and the existence of fine "nets" on the set of pure states. The second method uses the bound given by Eq. (3.52).

Method 1

A net of states is defined as follows: If $\epsilon > 0$ and g is a metric on $\mathbb{CP}^{d-1}$, an (ϵ, g)-net is defined to be a finite set of states $\mathcal{N}_{(\epsilon,g)} \subset \mathbb{CP}^{d-1}$ such that for any $|\psi\rangle \in \mathbb{CP}^{d-1}$ there exists $|\phi\rangle \in \mathcal{N}_{(\epsilon,g)}$ satisfying

$$g\left(|\psi\rangle, |\phi\rangle\right) \leq \epsilon.$$

It has been shown [101] that for $\epsilon \in (0, 1)$ and g induced by the 1-norm there exists an $(\epsilon, \| \|_1)$-net such that

$$\left|\mathcal{N}_{(\epsilon,\| \|_1)}\right| \leq \left(\frac{5}{\epsilon}\right)^{2d}. \tag{3.56}$$

This particular net is also shown to be a $(\frac{\epsilon}{2}, \| \|_2)$-net.

Let $\epsilon > 0$ and put $\| \|_2$ on $\mathbb{CP}^{d-1}$. From above, there exists an $\mathcal{N}_{(\epsilon,\| \|_2)}$ net of size $\left(\frac{5}{2\epsilon}\right)^{2d}$ on $\mathbb{CP}^{d-1}$. Suppose the minimum of the gate fidelity over $\mathbb{CP}^{d-1}$ occurs at $|\psi\rangle$. By definition there exists a state $|\phi\rangle \in \mathcal{N}_{(\epsilon,\| \|_2)}$ such that

$$\| |\psi\rangle - |\phi\rangle \|_2 \leq \epsilon.$$

Using the Lipschitz condition in Eq. (3.51), if Λ is a quantum operation,

$$\left|\mathcal{F}_{\Lambda,\mathcal{I}}\left(|\psi\rangle\right) - \mathcal{F}_{\Lambda,\mathcal{I}}\left(|\phi\rangle\right)\right| \leq 3\sqrt{2}\| |\psi\rangle - |\phi\rangle \|_2,$$

which implies

$$\mathcal{F}_{\Lambda,\mathcal{I}}\left(|\phi\rangle\right) - 3\sqrt{2}\epsilon \leq \mathcal{F}_{\Lambda,\mathcal{I}}\left(|\psi\rangle\right). \tag{3.57}$$

Therefore the minimum of $\mathcal{F}_{\Lambda,\mathcal{I}}$ over $\mathbb{CP}^{d-1}$ is bounded below by $\mathcal{F}_{\Lambda,\mathcal{I}}\left(|\phi\rangle\right) - 3\sqrt{2}\epsilon$ and the minimum over the net is a good approximation to the minimum over the entire space when ϵ is small.

As mentioned previously, by a simple concavity argument, the minimum of the gate fidelity over all mixed input states occurs at a pure state. Therefore Eq. (3.57) provides an estimate for the minimum over all mixed states. With the bound on the size of $\mathcal{N}_{(\epsilon,\| \|_2)}$ given in Eq. (3.56), this method will only be useful for small quantum systems. More

scalable bounds on the size of the net would imply the applicability of this method for larger quantum systems.

Property 2 from [56] (see Sec. A.4.2) is that a useful distance measure should be easy to calculate. The minimum gate fidelity has the drawback of not being easy to calculate analytically, even when a description of the noise process is available. However, convex optimization techniques can be used to numerically evaluate an estimate for the minimum when the noise process is known. The above lower bound implies that if one has a description of the noise then evaluating the minimum fidelity over a finite set of states gives an approximation of the minimum over all mixed quantum states. Tightening the bounds on the size of the net required would make this method more applicable.

This method also gives a clear experimental procedure for estimating the minimum gate fidelity (property 3 from [56]) without requiring process tomography. The idea is to be able to prepare a suitable net of states and determine the minimum fidelity over these states by performing measurements in the appropriate bases. Again, this minimum provides a good approximation to the minimum over all states but the obvious drawback is that the number of states scales poorly with the dimension of the system.

Method 2

The second method for estimating the minimum gate fidelity uses the concentration result for the gate fidelity given in Eq. (3.52). Let $Q > 0$ be fixed and suppose one is only interested in finding the smallest value $\mathcal{F}_{\Lambda,\mathcal{I}}$ can take such that any state $|\phi\rangle$ producing a smaller value lies in a set whose measure equals Q. In this context the smallest value is called the effective minimum, denoted $\mathcal{F}_{\text{eff}}$, given the tolerance Q. This problem is equivalent to finding the maximum over all $b \in [0,1]$ satisfying,

$$\mathbb{P}_{\mu_F}\left[\mathcal{F}_{\Lambda,\mathcal{I}} \in [0,b]\right] \leq Q.$$

The maximum value of b is equal to $\mathcal{F}_{\text{eff}}$ and depends on both d and Q.

By Eq. (3.52) for every $\epsilon > 0$,

$$\mathbb{P}_{\mu_F}\left[\mathcal{F}_{\Lambda,\mathcal{I}} \in \left[0, \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon\right]\right] \leq 2\text{exp}\left(\frac{-d\epsilon^2}{81\pi^3\ln(2)}\right).$$

This inequality can be used to find a non-trivial lower bound for $\mathcal{F}_{\text{eff}}$. Let $\epsilon_{Q,d}$ be the value of ϵ obtained when $Q = 2\text{exp}\left(\frac{-d\epsilon^2}{81\pi^3\ln(2)}\right)$,

$$\epsilon_{Q,d} = \sqrt{\frac{81\pi^3 \ln(2) \ln\left(\frac{2}{Q}\right)}{d}}.$$

By construction $\epsilon_{Q,d}$ satisfies $\mathbb{P}_{\mu_F}\left[\mathcal{F}_{\Lambda,\mathcal{I}} \in \left[0, \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon_{Q,d}\right]\right] \leq Q$ and so by definition,

$$\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} \geq \mathcal{F}_{\text{eff}} \geq \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \epsilon_{Q,d} = \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} - \sqrt{\frac{81\pi^3 \ln(2) \ln\left(\frac{2}{Q}\right)}{d}}. \tag{3.58}$$

This lower bound on $\mathcal{F}_{\text{eff}}$ is non-trivial since for fixed Q, $\epsilon_{Q,d} \to 0$ as $d \to \infty$. Therefore $\mathcal{F}_{\text{eff}} \to \overline{\mathcal{F}_{\Lambda,\mathcal{I}}}$ as $d \to \infty$, and the effective minimum and average of the gate fidelity become indistinguishable for large d.

3.7 Conclusion

In this chapter we have discussed and proven various properties of the quantum gate fidelity. We have shown the gate fidelity is not unique in that if $\dim(\mathcal{H}) = d \geq 4$ and $\mathcal{Q}$ is a quantum operation on $L(\mathcal{H})$ with a positive-definite Choi matrix, then there exists a channel $\mathcal{R} \neq \mathcal{Q}^\dagger$ (and $\mathcal{R} \neq \mathcal{Q}$) such that

$$\mathcal{F}_{\mathcal{Q},\mathcal{I}} = \mathcal{F}_{\mathcal{R},\mathcal{I}}.$$

A corollary of this result is that when $d \geq 4$ and $\mathcal{Q}$ is a depolarizing channel, there exist non-depolarizing channels $\mathcal{R}$ which produces the *same* gate fidelity as $\mathcal{Q}$. Since in this case $\mathcal{Q}$ has a constant gate fidelity on $\mathbb{CP}^{d-1}$, there exist non-depolarizing channels with a constant gate fidelity on $\mathbb{CP}^{d-1}$.

We have provided a method for calculating all moments of the gate fidelity $\mathcal{F}_{\mathcal{E},\mathcal{U}}$ between a unitary $\mathcal{U}$ and a quantum operation $\mathcal{E}$. Using this method we have obtained a closed form expression for $\text{Var}(\mathcal{F}_{\mathcal{E},\mathcal{U}})$ in terms of the Choi representation for $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$. A simple expression for the variance was given in the single qubit case and an explicit upper-bound for the variance was obtained for all d (see Eq. (3.37)). This upper-bound shows that for large quantum systems the variance scales as $O\left(\frac{1}{d}\right)$ for any $\mathcal{E}$ and $\mathcal{U}$.

Using Levy's lemma, an upper bound on the probability that a randomly chosen pure state produces a gate fidelity value far from the average was derived (see Eq. (3.52)). For

fixed ϵ, the upper bound converges to 0 exponentially quickly in the number of qubits comprising the quantum system. Eq.'s (3.37) and (3.52) provide a means for quantifying the fact that *all* quantum channels appear depolarizing with respect to the gate fidelity when d grows large. Expressions for this convergence are contained in Eq.'s (3.55) and (3.56). Lastly, we have provided a means for estimating the minimum gate fidelity in terms of nets of states on $\mathbb{CP}^{d-1}$ (Eq. (3.57)) and defining an "effective minimum" for the gate fidelity (Eq. (3.58)).

3.8 Discussion

Intuitively, the fact that Theorem 2 holds in higher dimensions is related to the complex geometry of the Bloch space representation of quantum states in higher dimensions. The simple Bloch sphere representation of a single qubit appears to indicate that Theorem 2 cannot be extended to $d = 2$. This was confirmed in Ref. [73] where the authors showed that for $d = 2$, two channels $\mathcal{Q}$ and $\mathcal{R}$ produce the same gate fidelity function if and only if their difference $\mathcal{R} - \mathcal{Q}$ is equal to the scaled difference of some unital quantum channel $\mathcal{E}$ and its dual $\mathcal{E}^\dagger$. The authors also showed the same is true for $d = 3$ but for the specific case of $\mathcal{Q}$ and $\mathcal{R}$ being unital. Thus the only case which remains open at this point is when $d = 3$ and at least one of $\mathcal{Q}$ or $\mathcal{R}$ is non-unital.

An entire family of open questions arising from theorem 2 relates to how two quantum channels which produce the same gate fidelity can differ with respect to a specific information-theoretic property. For instance, an interesting direction of research would be to analyze the extent to which two quantum channels which produce the same gate fidelity can differ in their capacities for transmitting information (see Sec. B.5 for a discussion of channel capacities). As well, since the diamond norm distance between channels has a well-defined operational meaning (see Sec. A.4.2) it would be interesting to analyze how far apart two channels with the same gate fidelity can be with respect to the diamond norm.

One of the central points of this thesis, and specifically Chapter 2, is the importance of estimating partial information about a noise process in a completely scalable manner. For instance [66] has discussed estimating the average fidelity based on classical fidelity bounds on complementary bases. More recently, twirling [14, 40] and randomization methods [46, 47, 84, 134, 99] have attempted to provide a scalable means for estimating the average fidelity as well as more detailed features of the noise. Methods for estimating the variance of the fidelity over the twirling/randomizing gate set will provide even more information about the unknown noise model.

While Eq. (3.37) appears to depend non-trivially on all elements of the χ-matrix it may be the case that subsets of terms in the expression correspond to quantities that have operational significance and can be obtained via some experimental procedure. If this were the case then it would be worthwhile to pursue scalable methods for estimating the variance of a quantum channel. Calculating the variance using the current expression however will clearly not be scalable in n. For small numbers of qubits it will still be worthwhile to perform process tomography to obtain a complete description of the noise and calculate the variance using Eq. (3.37).

In [84] it is suggested that the variance of the fidelity measured under the proposed randomized benchmarking protocol may provide useful information about the extent to which the noise is coherent (understood here to mean the noise does not consist solely of Pauli errors). While this may be the case for a small number of qubits n, we have shown that the variance of the gate fidelity will decrease exponentially quickly in n. This implies that an exponentially increasing number of repetitions of the protocol would be required to obtain information about the coherence of the noise, making the method infeasible for even moderately large systems. Moreover our expression for the variance shows that it depends in a non-trivial way on both the diagonal and off-diagonal elements of the χ-matrix. Hence the extent of the coherence of the noise model can not be inferred from an estimate of the variance alone.

Also note that, assuming the noise is effectively independent of the gate set, in order for the variance to be independent of the initial state and the particular choice of randomizing gates, the randomizing gates must comprise a unitary 4-design. Of course using Haar-random gates will produce a variance that depends only on the noise model, however such a protocol is practical for a small number of qubits since implementing Haar-random unitaries is exponentially hard in n. Recently, the existence of efficient approximate unitary 4-designs has been proven [63] and randomizing under such a gate set may provide methods for estimating the variance of the gate fidelity.

Concentration of measure techniques, and specifically Levy's lemma, have played an extremely important role in quantum information in recent years. The concentration of measure result we have obtained for the gate fidelity is just one example of generically quantifying the large-dimensional behaviour of quantum systems and operations. For instance concentration of measure has been utilized to prove the existence of subspaces of bipartite quantum systems consisting entirely of entangled states [65], explain thermalization in statistical mechanics [116], and construct counter-examples to the additivity conjecture [64]. An extremely interesting direction of further research is to understand more generic, large-dimensional features of quantum systems by employing these concentration techniques.

The minimum gate fidelity is a stronger characterization of a noise process than the average and thus experimental methods for estimating it are a useful direction of further research. Some methods for calculating the minimum gate fidelity given a description of the noise process are given in Ref.'s [73, 87]. Ref. [87] casts the minimum gate fidelity of a quantum channel in terms of the $S(1)$-norm [74] of the compression of the Choi matrix of the channel to the symmetric subspace. Efficient methods for computing the minimum in the single qubit case via a semidefinite program are also given. Ref. [87] represents the minimum gate fidelity in terms of numerical ranges and provides a method for computing the minimum in the case that the channel is unitary.

APPENDICES

Appendix A

Quantum Mechanics

A.1 State Space of a Quantum System

This thesis will deal only with finite-dimensional quantum systems, therefore quantum systems will be represented by a complex Hilbert space $\mathcal{H}$ of dimension $d < \infty$. The standard isomorphism between $\mathcal{H}$ and $\mathbb{C}^d$ will be assumed without mention throughout the presentation. Denote the set of linear operators on $\mathcal{H}$ by $L(\mathcal{H})$. The set of pure states for the system is represented by $\mathbb{C}^d$ modulo phase factors, ie. complex projective space $\mathbb{CP}^{d-1}$. Equivalently, these are elements ρ of $L(\mathcal{H})$ that satisfy

$$\rho^\dagger = \rho \text{ and } \rho^2 = \rho. \tag{A.1}$$

Thus they can be associated with the set of rank 1 projectors. Hence the pure state $|\psi\rangle\langle\psi|$ is a representative of the equivalence class of vectors $e^{i\theta}|\psi\rangle$ in $\mathcal{H}$. More generally, mixed states for the system are described by the set of positive trace-1 operators in $L(\mathcal{H})$, and will be denoted by $\mathcal{D}(\mathcal{H})$. Elements of $\mathcal{H}$ will be labeled using Dirac notation. A vector in the Hilbert space is written in "ket" form $|\psi\rangle$ and the inner product of $|\psi\rangle$ with $|\phi\rangle$ is written $\langle\phi||\psi\rangle$, or more simply, $\langle\phi|\psi\rangle$. The object $\langle\phi|$ is called a "bra" and represents the unique linear functional f_ϕ on $\mathcal{H}$ given by

$$f_\phi(|\psi\rangle) = \langle\phi|\psi\rangle. \tag{A.2}$$

The outer product of $|\psi\rangle$ and $|\phi\rangle$ is denoted $|\phi\rangle\langle\psi|$ and by linearity of the inner product is a linear operator on $\mathcal{H}$. $L(\mathcal{H})$ can be made into a Hilbert space by defining the inner

product of σ with τ to be

$$\langle \tau | \sigma \rangle = tr(\tau^\dagger \sigma) \tag{A.3}$$

where $\tau^\dagger$ is the adjoint of τ. This inner product is called the trace inner product and, unless otherwise stated, $L(\mathcal{H})$ will be assumed to have this inner product defined on it.

A.1.1 Composite Quantum Systems

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be finite dimensional Hilbert spaces with bases $\{|v_i\rangle\}_{i=1}^n$ and $\{|w_j\rangle\}_{j=1}^m$ respectively. The direct sum of these two spaces has basis given by the union of the bases of its component spaces $\{|v_1\rangle, ..., |v_n\rangle, |w_1\rangle, ..., |w_m\rangle\}$. The tensor product of $\mathcal{H}_1$ and $\mathcal{H}_2$ has basis $\{|v_1\rangle \otimes |w_1\rangle, ..., |v_1\rangle \otimes |w_m\rangle, ..., |v_n\rangle \otimes |w_1\rangle, ..., |v_n\rangle \otimes |w_m\rangle\}$.

Individual state spaces of n particles combine classically through the direct sum while quantum states combine through the tensor product. Thus, the dimension of the state space of multiple classical particles grows linearly with the number of particles, since $\dim(\mathcal{H}_1 \oplus \mathcal{H}_2) = \dim(\mathcal{H}_1) + \dim(\mathcal{H}_2)$. In the quantum case, the dimension of the composite system increases as $\dim(\mathcal{H}_1)\dim(\mathcal{H}_2)$. The extension to multi-partite quantum systems is performed by taking the tensor product of the spaces describing each party. We will sometimes denote the t-fold tensor product of a state $|\psi\rangle$ with itself by the expression $|\psi\rangle^{\otimes t}$.

A state $\rho \in L(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is called separable if it can be written in the form

$$\rho = \sum_{i=1}^{r} p_i \left(\sigma_i \otimes \tau_i\right) \tag{A.4}$$

where $\{p_i\}$ is a probability distribution. States that are not separable are called entangled and are of great importance in quantum information processing. Interestingly, pure entangled states are dense in the space of all pure quantum states [111], which means that if an arbitrary state is chosen from the state space then any open set containing this state will also contain an entangled state. For a Hilbert space $\mathcal{H}$ of dimension d and orthonormal basis $\{|i\rangle\}$ we will make extensive use of the "maximally entangled" state $|\psi_0\rangle$ defined by

$$
\begin{aligned}
|\psi_0\rangle &:= \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle \otimes |j\rangle, \\
\chi_0 &:= |\psi_0\rangle\langle\psi_0|.
\end{aligned} \tag{A.5}
$$

A.2 Evolution of Quantum Systems: CPTP maps and Useful Representations

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ represent finite-dimensional quantum systems of dimensions d_1 and d_2 respectively. The set of linear superoperators from $L(\mathcal{H}_1)$ to $L(\mathcal{H}_2)$ will be denoted by $\mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$. A quantum channel, or quantum operation, $\mathcal{E}$ is a completely positive, trace-preserving mapping from $L(\mathcal{H}_1)$ into $L(\mathcal{H}_2)$ [108]. The set of quantum channels contained in $\mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ will be denoted by $\mathcal{S}(\mathcal{H}_1, \mathcal{H}_2)$. Quantum channels describe how an input quantum system is changed under some process or time-evolution. Note that in general the output system of the evolution will be described by a different Hilbert space than the input. In the case that $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$, $\mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ will be denoted $\mathcal{T}(\mathcal{H})$ and similarly for $\mathcal{S}(\mathcal{H}_1, \mathcal{H}_2)$. An important sub-class of quantum channels is the set of "unital" channels, ie. those which map $\mathbb{1}_{\mathcal{H}_1}$ to $\mathbb{1}_{\mathcal{H}_2}$. Unitary channels are unital and describe the dynamics of a quantum system that is isolated from environmental interactions. Given a linear superoperator $\mathcal{E} \in \mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$, the adjoint, or dual, map $\mathcal{E}^\dagger \in \mathcal{T}(\mathcal{H}_2, \mathcal{H}_1)$ is uniquely defined in the usual manner by the equation

$$
tr(\mathcal{E}^\dagger(\sigma)\tau) = tr(\sigma^\dagger \mathcal{E}(\tau)). \tag{A.6}
$$

Two important quantum operations on bipartite matrices are the partial trace and partial transpose. For $A \in L(\mathcal{H} \otimes \mathcal{H})$ these operations are denoted and defined as follows:

- The *partial trace* over one subsystem: We will denote the partial trace over subsystem i by $\mathrm{tr}_i[A]$. When the partial trace is applied to a state, it generates the reduced density matrix of the remaining subsystem, for instance $\rho_2 = \mathrm{tr}_1\rho$.

  Explicit expressions for the partial trace operation of an operator $A = A_1 \otimes A_2$ are given by:

$$
\begin{aligned}
\mathrm{tr}_1[A] &= \mathrm{tr}(A_1)A_2, \\
\mathrm{tr}_2[A] &= \mathrm{tr}(A_2)A_1.
\end{aligned}
$$

This definition is extended to any element of $L(\mathcal{H} \otimes \mathcal{H})$ by linearity.

- The *partial transpose* over one subsystem: For any bipartite matrix A, we will denote the partial transpose of A with respect to the ith subsystem by A^{T_i} (similarly, A^T indicate the full transpose of A, with respect to both subsystems). Partial transposition is *not* a completely positive operation; in particular, it transforms many entangled states into negative matrices. Explicit expressions for the partial transpose operation on the subsystems are given by:

$$\begin{aligned} A^{T_1} &= \sum_{k,l=0}^{d-1} \left(|k\rangle\langle l| \otimes \mathbb{1}\right) A \left(|k\rangle\langle l| \otimes \mathbb{1}\right) \\ A^{T_2} &= \sum_{k,l=0}^{d-1} \left(\mathbb{1} \otimes |k\rangle\langle l|\right) A \left(\mathbb{1} \otimes |k\rangle\langle l|\right). \end{aligned}$$

There are many ways to represent a completely positive, trace-preserving mapping which include the standard representation, Choi matrix representation [29], Kraus representation [29, 85], Stinespring's representation [137] and the χ-matrix representation. A good reference for completely positive maps and their representations is given by [112]. We briefly describe these representations as they will be used frequently throughout the presentation. We also show that the Choi and χ-representations can be identified by choosing appropriate bases to write the respective representations in.

Let $\Lambda \in \mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ and $\{Q_i\}$, $\{R_j\}$ be bases for $L(\mathcal{H}_1)$ and $L(\mathcal{H}_2)$ respectively. The standard representation of Λ with respect to the above bases is the d_2^2 by d_1^2 matrix,

$$\Lambda_{i,j} = \mathrm{tr}\left(R_i^\dagger \Lambda(Q_j)\right).$$

While this representation is both natural and useful, it is basis-dependent and complete positivity is not easily tested in this representation.

The Choi representation of a linear superoperator Λ on $L(\mathcal{H}_1)$, denoted $C(\Lambda)$, is the linear operator on $\mathcal{H}_2 \otimes \mathcal{H}_1$ given by,

$$C(\Lambda) = \sum_{(a,b) \in \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_1}} \Lambda(|a\rangle\langle b|) \otimes |a\rangle\langle b| = (\Lambda \otimes \mathcal{I})\left(d_1 \chi_0\right) \tag{A.7}$$

where we recall χ_0 is the projector onto the maximally entangled Bell state $|\psi_0\rangle$,

$$\chi_0 = |\psi_0\rangle\langle\psi_0| = \left(\frac{1}{\sqrt{d_1}}\sum_{a=1}^{d_1}|a\rangle\otimes|a\rangle\right)\left(\frac{1}{\sqrt{d_1}}\sum_{b=1}^{d_1}\langle b|\otimes\langle b|\right). \tag{A.8}$$

Note that the Choi representation of a linear superoperator is unique and the association $\Lambda \to C(\Lambda)$ is a linear isomorphism between $\mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ and $L(\mathcal{H}_2 \otimes \mathcal{H}_1)$. Note also that for any Λ_1 and Λ_2, $C(\Lambda_1 \otimes \Lambda_2) = C(\Lambda_1) \otimes C(\Lambda_2)$. From equation (A.7), Λ is completely positive and trace-preserving if and only if $\frac{1}{d_1}C(\Lambda)$ is a quantum state in $L(\mathcal{H}_2 \otimes \mathcal{H}_1)$. Therefore the mapping $\Lambda \to \frac{1}{d_1}C(\Lambda)$ is a linear isomorphism between quantum operations and quantum states. The state $J(\Lambda) := \frac{1}{d_1}C(\Lambda)$ is commonly called the *Jamiolkowski state* associated to Λ and the isomorphism is known as the *Choi-Jamiolkowski isomorphism*.

When $C(\Lambda)$ is written with respect to the basis $\{|a\rangle|b\rangle\langle c|\langle d|\}$ (where a and c range from 0 to $d_2 - 1$, b and d range from 0 to $d_1 - 1$, and we assume the right-most index varies fastest in tensor product state bases) the resulting matrix is called the *Choi matrix*. The *Jamiolkowski matrix* is naturally defined as the Choi matrix multiplied by $\frac{1}{d_1}$. Note that this definition does not imply the Choi matrix corresponds to simply block-constructing a matrix via $(\Lambda(|i\rangle\langle j|))_{i,j}$. This correspondence would hold however if we either assumed that the left-most index varies fastest in tensor product state bases or defined $C(\Lambda) = (\mathcal{I} \otimes \Lambda)(d_1\chi_0)$.

A "Kraus representation" of the linear superoperator Λ can be obtained from $C(\Lambda)$. By the singular value decomposition,

$$C(\Lambda) = \sum_{i=1}^{k}|a_i\rangle\langle b_i|$$

where the $|a_i\rangle$ and $|b_i\rangle$ are proportional to the left and right singular vectors of $J(\Lambda)$ respectively, and k is the rank of $J(\Lambda)$. There is an obvious inner-product isomorphism between $L(\mathcal{H}_1, \mathcal{H}_2)$ with the Hilbert-Schmidt inner product and $\mathcal{H}_2 \otimes \mathcal{H}_1$ with the standard inner product, defined by $|a\rangle\langle b| \to \mathrm{vec}(|a\rangle\langle b|) = |a\rangle \otimes |b\rangle$. If A_i and B_i are the unique linear operators in $L(\mathcal{H}_1, \mathcal{H}_2)$ satisfying $\mathrm{vec}(A_i) = |a_i\rangle$ and $\mathrm{vec}(B_i) = |b_i\rangle$ respectively, then for every $M \in L(\mathcal{H}_1)$,

$$\Lambda(M) = \sum_{i=1}^{k} A_i M B_i^\dagger. \tag{A.9}$$

The above expression is called a Kraus representation for Λ and the set of $\{A_i, B_i\}$ are called Kraus operators for the linear superoperator. If Λ is completely positive and trace preserving (ie. a quantum operation) then $B_i = A_i$ for each i and $\sum_{i=1}^{k} A_i^\dagger A_i = \mathbb{1}_{\mathcal{H}_1}$. Unlike the Choi representation, a Kraus representation is not unique in that there is a unitary freedom in the Kraus operators describing a quantum channel Λ [108]. If the set of operators $\{E_j\}$ is defined by

$$E_j = \sum_l U_{jl} A_l \tag{A.10}$$

for some unitary matrix U, then $\{E_j\}$ describes the same quantum operation as the $\{A_i\}$. The converse is also true: if $\{E_j\}$ and $\{A_i\}$ define the same quantum operation then they are unitarily related as above, where the cardinality of the index sets for $\{E_j\}$ and $\{A_i\}$ can be made equal by appending the necessary number of zero operators to the smaller set.

Note that unital quantum channels have Kraus operators $\{A_i\}$ that satisfy the additional constraint

$$\sum_i A_i A_i^\dagger = \mathbb{1}_{\mathcal{H}_2} \tag{A.11}$$

and the set of Kraus operators for unitary evolution consists of a single unitary operator.

If Λ is completely positive with Kraus representation given by $\{A_i\}$ then it is not hard to see that the adjoint map is completely positive with Kraus operators $\{A_i^\dagger\}$. Indeed, we need only verify that the equation defining the adjoint map of Λ is satisfied using the set of Kraus operators $\{A_i^\dagger\}$. By the linearity and cyclic properties of the trace we have,

$$\mathrm{tr}\left(A_i^\dagger \sigma A_i \tau\right) = \mathrm{tr}\left(A_i \tau A_i^\dagger \sigma\right). \tag{A.12}$$

A Stinespring representation for $\Lambda \in \mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ can be constructed from a Kraus representation for Λ. Let $\mathcal{H}_3$ have dimension equal to $k = \mathrm{rank}(J(\Lambda))$ and $\{|e_i\rangle\}_{i=1}^k$ be an orthonormal basis for $\mathcal{H}_3$. Then there exists linear operators V and W taking $\mathcal{H}_1$ to $\mathcal{H}_2 \otimes \mathcal{H}_3$ defined by $V = \sum_{i=1}^k A_i \otimes |e_i\rangle$ and $W = \sum_{i=1}^k B_i \otimes |e_i\rangle$ such that for every $M \in L(\mathcal{H}_1)$,

$$\Lambda(M) = \mathrm{tr}_{\mathcal{H}_3} V M W^\dagger. \tag{A.13}$$

Here "$\mathrm{tr}_{\mathcal{H}_3}$" denotes the partial trace operation with respect to $\mathcal{H}_3$. The above expression is called a Stinespring representation and, like the Kraus representation, is not unique. If Λ is a quantum operation then $V = W$ and $V^\dagger V = \mathbb{1}_{\mathcal{H}_1}$ which implies that V is an isometry.

Lastly, a useful representation in quantum process tomography is the χ-representation of a quantum operation. If the linear superoperator Λ has Kraus operators $\{A_i, B_i\} \in L(\mathcal{H}_1, \mathcal{H}_2)$, and if $\{Q_j\}$ is a basis for $L(\mathcal{H}_1, \mathcal{H}_2)$, we can expand the Kraus operators in this basis and write the action of Λ on $M \in L(\mathcal{H}_1)$ as,

$$\begin{aligned} \Lambda(M) &= \sum_{i=1}^{k} A_i M B_i^\dagger \\ &= \sum_{i,j} \chi_{i,j} Q_i M Q_j^\dagger. \end{aligned}$$

The $d_1 d_2$ by $d_1 d_2$ matrix $\chi_{i,j}$ is called the χ-matrix for Λ and is unique given the choice of basis $\{Q_i\}$ (it does not depend on the choice of Kraus operators for Λ). We show next that $\chi_{i,j}$ written in the basis $\{Q_i\}$ can be identified with the Jamiolkowski representation written in a bipartite basis determined by the $\{Q_i\}$.

The χ and Jamiolkowski representations can be identified in the following manner: If $\Lambda \in \mathcal{T}(\mathcal{H}_1, \mathcal{H}_2)$ then,

$$\begin{aligned} J(\Lambda) &= \frac{1}{d} \sum_{a,b} \Lambda(|a\rangle\langle b|) \otimes |a\rangle\langle b| \\ &= \frac{1}{d} \sum_{a,b} \left(\sum_{i,j} \chi_{i,j} Q_i |a\rangle\langle b| Q_j^\dagger \right) \otimes |a\rangle\langle b| \\ &= \sum_{i,j} \chi_{i,j} \left(Q_i \otimes \mathbb{1}\right) |\psi_0\rangle\langle\psi_0| \left(Q_j^\dagger \otimes \mathbb{1}\right) \end{aligned}$$

where we claim that $\{(Q_i \otimes \mathbb{1})|\psi_0\rangle\}_{i=0}^{d_1 d_2 - 1}$ is a basis for the bipartite space $\mathcal{H}_2 \otimes \mathcal{H}_1$ (proven below). Hence the χ-matrix of Λ relative to $\{Q_i\}$ is *equal* to the Jamiolkowski state of Λ written with respect to the basis $\{(Q_i \otimes \mathbb{1})|\Psi\rangle\}$. Therefore there is no less of generality in writing χ to represent the linear operator $J(\Lambda)$. Hence, throughout the rest of the presentation, "$\chi_{i,j}$" will (unambiguously) refer to either $J(\Lambda)$ written in the bipartite basis $\{(Q_i \otimes \mathbb{1})|\Psi\rangle\}$ or the χ-matrix of Λ with respect to $\{Q_i\}$.

Proposition 1. *If* $\{Q_i\}$ *is a basis for* $L(\mathcal{H}_1, \mathcal{H}_2)$ *then* $\{(Q_i \otimes \mathbb{1})|\psi_0\rangle\}_{i=0}^{d_1 d_2 - 1}$ *is a basis for the bipartite space* $\mathcal{H}_2 \otimes \mathcal{H}_1$.

Proof. Since the Q_i are linearly independent, if

$$\sum_{i=0}^{d_1 d_2 - 1} \lambda_i Q_i = 0 \tag{A.14}$$

for scalars λ_i it must be that $\forall i$, $\lambda_i = 0$. Now suppose that for $d_1 d_2$ scalars λ_i,

$$\sum_{i=0}^{d_1 d_2 - 1} \lambda_i (Q_i \otimes \mathbb{1})|\psi_0\rangle = 0. \tag{A.15}$$

Then since,

$$\begin{aligned}
\sum_{i=0}^{d_1 d_2 - 1} \lambda_i (Q_i \otimes \mathbb{1})|\psi_0\rangle &= \frac{1}{\sqrt{d}} \sum_{a=0}^{d_1 - 1} \sum_{i=0}^{d_1 d_2 - 1} \lambda_i (Q_i \otimes \mathbb{1})|a\rangle \otimes |a\rangle \\
&= \frac{1}{\sqrt{d}} \sum_{a=0}^{d_1 - 1} \left(\sum_{i=0}^{d_1 d_2 - 1} \lambda_i Q_i |a\rangle \right) \otimes |a\rangle \\
&:= \frac{1}{\sqrt{d}} \sum_{a=0}^{d_1 - 1} |\tilde{a}\rangle \otimes |a\rangle
\end{aligned} \tag{A.16}$$

we have

$$\frac{1}{\sqrt{d}} \sum_{a=0}^{d_1 - 1} |\tilde{a}\rangle \otimes |a\rangle = 0. \tag{A.17}$$

Now since for any set of d_1 non-zero vectors $|\{\phi_i\rangle\}_{i=0}^{d_1 - 1}$, the set $\{|\phi_i\rangle \otimes |i\rangle\}_{i=0}^{d_1 - 1}$ is linearly independent we get that the set of non-zero vectors contained in $\{|\tilde{a}\rangle \otimes |a\rangle\}_{a=0}^{d_1 - 1}$ is linearly

independent (a subset of a linearly independent set is itself obviously linearly independent). Thus, if Eq. (A.17) holds it must be the case that each $|\tilde{a}\rangle \otimes |a\rangle$ is equal to 0, ie. $|\tilde{a}\rangle$ is the zero vector for each a. Hence by the definition of the $|\tilde{a}\rangle$, for every a,

$$\sum_{i=0}^{d_1 d_2 - 1} \lambda_i Q_i |a\rangle = 0 \tag{A.18}$$

and so for every $a \in \{0, ...d_1 - 1\}$, $|a\rangle$ is in the nullspace of the operator $\sum_{i=0}^{d_1 d_2 - 1} \lambda_i Q_i$. Since this operator is in $L(\mathcal{H}_1, \mathcal{H}_2)$ and $\{|a\rangle : a \in \{0, ...d_1 - 1\}\}$ is a basis for $\mathcal{H}_1$ this implies $\sum_{i=0}^{d_1 d_2 - 1} \lambda_i Q_i$ maps all of $\mathcal{H}_1$ to the zero vector. Hence

$$\sum_{i=0}^{d_1 d_2 - 1} \lambda_i Q_i = 0 \tag{A.19}$$

and by the assumption of the Q_i being linearly independent we get for every $i \in \{0, ..., d_1 d_2 - 1\}$, $\lambda_i = 0$. Thus, $\{(Q_i \otimes \mathbb{1}) |\psi_0\rangle\}_{i=0}^{d_1 d_2 - 1}$ is a basis for the bipartite space $\mathcal{H}_2 \otimes \mathcal{H}_1$.

□

Note that for a quantum operation $\mathcal{E}$, even though $J(\mathcal{E})$ as defined can be associated to a quantum state, writing $J(\mathcal{E})$ with respect to $\{(Q_i \otimes \mathbb{1}) |\Psi\rangle\}$ may produce a positive semidefinite matrix $\chi_{i,j}$ that does not have unit trace. It is straightforward to show however that if $\{Q_i\}$ is an orthogonal basis of $L(\mathcal{H})$ normalized so that $\mathrm{tr}\left(Q_j^\dagger Q_i\right) = \delta_{i,j} d$, then $\chi_{i,j}$ is a positive semi-definite, trace-1 matrix. A standard example of such a basis $\{Q_i\}$ is the set of (normalized) matrix units $\{\sqrt{d}|k\rangle\langle l|\}$, $k, l \in \mathbb{Z}_d$. Later we discuss other bases satisfying these conditions which will be more convenient for the calculations we deal with in this thesis.

A particular quantum operation that is of significant interest in quantum information theory and is used extensively in this thesis is the depolarizing channel. Depolarizing channels on $L\left(\mathbb{C}^d\right)$ are convex combinations of the identity mapping $\mathcal{I}$ and the "totally depolarizing" mapping Ω given by

$$\Omega(X) = \mathrm{tr}(X) \frac{\mathbb{1}}{d}.$$

Restricting the domain to quantum states implies that a depolarizing channel Φ has the form,

$$\Phi(\rho) = p\rho + (1-p)\frac{\mathbb{1}}{d}$$

where $p \in [0,1]$ and ρ is an arbitrary quantum state. Clearly $p = 1$ corresponds to the identity map $\mathcal{I}$ and $p = 0$ corresponds to Ω. The set of depolarizing channels in $\mathcal{S}(\mathcal{H})$ will be denoted by $\mathcal{R}(\mathcal{H})$.

Sets of Kraus operators for the totally depolarizing channel are given by any unitary 1-design (see Sec. C.1 or [40]), examples of which are the generalized Gell-Mann basis [54], the Heisenberg-Weyl basis and, when $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, the n-fold tensor product of single qubit Pauli operators (see Sec. B.1.1). For an excellent discussion of these bases and depolarizing channels see [26]. Note that all of these bases contain $\mathbb{1}$ with the remaining operators being traceless. Let $\left\{\frac{P_i}{d} : i \in \{0, ..., d^2-1\}\right\}$ represent any one of these orthonormal bases with $P_0 = \frac{\mathbb{1}}{d}$. Then,

$$\frac{1}{d^2}\sum_{i=0}^{d^2-1} P_i \rho P_i^\dagger = \frac{\mathbb{1}}{d}$$

which gives,

$$\begin{aligned}\Phi(\rho) &= p\rho + \frac{1-p}{d^2}\sum_i P_i \rho P_i^\dagger \\ &= \left(p + \frac{1-p}{d^2}\right)\rho + \frac{1-p}{d^2}\sum_{i=1}^{d^2} P_i \rho P_i^\dagger .\end{aligned}$$

Therefore the Kraus operators for Φ are $\left(\sqrt{p + \frac{1-p}{d^2}}\right)\mathbb{1}$ and $\left\{\frac{\sqrt{1-p}}{d} P_i : i \in \{1, ..., d^2-1\}\right\}$.

A.3 Measurement

Measurement of a quantum system obeys different transformation rules than those described above. A measurement is described by a set of linear operators $\{M_m\}$ satisfying the completeness relation

$$\sum_m M_m^\dagger M_m = \mathbb{1}. \tag{A.20}$$

The measurement operators are indexed by the measurement outcomes m. If the state of the system is ρ, then the probability of obtaining outcome m is given by

$$p(m) = tr(M_m^\dagger M_m \rho). \tag{A.21}$$

By the completeness relation above, $p(m)$ is a normalized probability distribution since

$$\begin{aligned} \sum_m p(m) &= \sum_m tr(M_m^\dagger M_m \rho) \\ &= tr(\sum_m M_m^\dagger M_m \rho) \\ &= tr(\rho) \\ &= 1. \end{aligned} \tag{A.22}$$

The state of the system after the measurement is

$$\frac{M_m \rho M_m^\dagger}{p(m)}. \tag{A.23}$$

In many cases it is convenient to define the positive operator

$$E_m = M_m^\dagger M_m \tag{A.24}$$

and make the necessary replacement in the above expressions. The set of operators $\{E_m\}$ is called a positive operator valued measure (POVM). POVM's are especially useful when only measurement statistics are of interest.

The simplest kind of measurement is when the measurement operators consist of positive-rank projection operators $P_m = \sum_j |\psi_m^j\rangle\langle\psi_m^j|$. Equivalently, the P_m are positive-rank Hermitian operators satisfying $P_m^2 = P_m$. Such a measurement is called a projective measurement. Clearly such measurements are in a 1-1 correspondence with Hermitian operators where the eigenvalues are the measurement outcomes m and projectors onto the associated

eigenspace are the projectors for the measurement. The probability of outcome m is given by

$$p(m) = tr(P_m \rho) \tag{A.25}$$

and the post-measurement state is

$$\frac{P_m \rho P_m}{p(m)}. \tag{A.26}$$

A.4 Distinguishing Quantum States and Operations

Many methods for distinguishing states and operations in quantum theory rely on the concept of a metric, or some quantity that resembles a metric. A metric is defined as follows,

Definition 3. *Metric*

Let X be a set. A metric on X is a function $d : X \to \mathbb{R}$ that satisfies the following properties:

1. $\forall\ x,\ y \in X,\ d(x, y) \geq 0$
2. $d(x, y) = 0$ *if and only if* $x = y$
3. $\forall\ x,y \in X,\ d(x, y) = d(y, x)$
4. $\forall\ x,y,z \in X,\ d(x, y) \leq d(x, z) + d(z, y)$.

We first discuss distinguishing quantum states and then use many of the results to discuss various useful distinguishability measures on quantum operations.

A.4.1 Distinguishing Quantum States

There are many methods for measuring the distance between, or distinguishing, quantum states (for a comprehensive discussion see [51]). Two important measures of distance are the trace distance, which is a metric, and the fidelity, which is not. Before defining these quantities, we look at their classical analogues.

Definition 4. *Classical Trace Distance and Fidelity*

Let $\{p_x\}$ and $\{q_x\}$ be two probability distributions over the same index set. The classical trace distance, $\mathcal{D}_C$, between the distributions is the l_1 distance between the distributions, that is,

$$\mathcal{D}_C = \sum_x |p_x - q_x| \,. \tag{A.27}$$

The classical fidelity, $\mathcal{F}_C$, between the distributions is

$$\mathcal{F}_C(p_i, q_i) = \sum_i \sqrt{p_i q_i}. \tag{A.28}$$

$\mathcal{F}_C$ is clearly not a metric since if $p_i = q_i$ for every i, $\mathcal{F}_C(p_i, q_i) = 1$. Hence, $\mathcal{F}_C(p_i, q_i)$ being near 1 indicates the probability distributions are close to each other. We now define the trace distance in the quantum case and discuss some of its properties.

Definition 5. *Trace Distance*

The trace distance $\mathcal{D}$ between two quantum states ρ and σ is equal to the metric induced by the trace inner product. Specifically,

$$\mathcal{D}(\rho, \sigma) = tr\,|\rho - \sigma| = \|\rho - \sigma\|_1 \tag{A.29}$$

where as usual for $A \in L(\mathcal{H})$, $|A| = \sqrt{A^\dagger A}$.

Clearly the trace distance is unitarily invariant, ie. for all unitaries U,

$$\mathcal{D}(\rho, \sigma) = \mathcal{D}(U\rho U^\dagger, U\sigma U^\dagger). \tag{A.30}$$

The following characterizes the trace distance in terms of measurement statistics [108].

Proposition 2. *Let ρ and σ be quantum states and $\{E_m\}$ be an arbitrary POVM. Define the probability distributions p_m and q_m by $p_m = tr(\rho E_m)$ and $q_m = tr(\rho E_m)$. Then*

$$\mathcal{D}(\rho, \sigma) = max_{\{E_m\}} \mathcal{D}_C(p_m, q_m) \tag{A.31}$$

where the maximization is over all POVM's $\{E_m\}$.

Hence the trace distance is the largest possible classical trace distance between the probability distributions arising from a POVM. An important property of the trace distance is that of strong convexity [108].

Proposition 3. *Let p_m and q_m be probability distributions over the same index set. Then, for states ρ_m and σ_m defined on this index set,*

$$\mathcal{D}\left(\sum_m p_m\rho_m, \sum_m q_m\sigma_m\right) \leq \mathcal{D}_C(p_m, q_m) + \sum_m p_m\mathcal{D}(\rho_m, \sigma_m). \tag{A.32}$$

In the special case of $p_m = q_m\ \forall m$ we get

$$\mathcal{D}\left(\sum_m p_m\rho_m, \sum_m p_m\sigma_m\right) \leq \sum_m p_m\mathcal{D}(\rho_m, \sigma_m). \tag{A.33}$$

Thus the trace distance is jointly convex in its inputs. Next, we define the fidelity and discuss some analogous properties to those for the trace distance.

Definition 6. *Fidelity*

The fidelity, $\mathcal{F}$, between ρ and σ is

$$\mathcal{F}(\rho, \sigma) = tr\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}. \tag{A.34}$$

The fidelity satisfies all of the properties of a metric except for being zero when $\rho = \sigma$. When ρ (or σ by symmetry) is a projector $|\psi\rangle\langle\psi|$ we have the following simple form for the fidelity

$$\begin{aligned}\mathcal{F}(\rho, \sigma) &= tr\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}} \\ &= tr\sqrt{\langle\psi|\sigma|\psi\rangle|\psi\rangle\langle\psi|} \\ &= \sqrt{\langle\psi|\sigma|\psi\rangle}.\end{aligned} \tag{A.35}$$

As with the trace distance, the fidelity is unitarily invariant. There is also an analogous characterization of the fidelity in terms of measurement statistics [108].

Proposition 4. *Let ρ and σ be quantum states and $\{E_m\}$ be an arbitrary POVM. Define the probability distributions p_m and q_m by $p_m = tr(\rho E_m)$ and $q_m = tr(\rho E_m)$. Then*

$$\mathcal{F}(\rho, \sigma) = min_{\{E_m\}} \mathcal{F}_C(p_m, q_m) \tag{A.36}$$

where the minimization is over all POVMs.

Analogous to the strong convexity result for the trace distance, the fidelity satisfies a strong concavity property.

Proposition 5. *Let p_m and q_m be probability distributions over the same index set. Then, for states ρ_m and σ_m defined on the same index set,*

$$\mathcal{F}\left(\sum_m p_m \rho_m, \sum_m q_m \sigma_m\right) \geq \sum_m \sqrt{p_m q_m} \mathcal{F}(\rho_m, \sigma_m). \tag{A.37}$$

In the special case of $p_m = q_m \, \forall m$ we get

$$\mathcal{F}\left(\sum_m p_m \rho_m, \sum_m p_m \sigma_m\right) \geq \sum_m p_m \mathcal{F}(\rho_m, \sigma_m). \tag{A.38}$$

Thus the fidelity is jointly concave in its inputs.

A.4.2 Distinguishing Quantum Operations

As with quantum states, there are many methods for measuring the distance between quantum operations (for a good discussion on various distance measures see [56]). Mirroring the path of the previous section regarding quantum states, we first focus on two important methods for distinguishing operations, the diamond norm distance and channel fidelity. These are natural extensions of the trace distance and fidelity on quantum states and so, not surprisingly, the diamond norm distance is a metric whereas the channel fidelity is not. Afterwards we discuss another class of distance measures which is particularly relevant in the perturbative expansion of randomized benchmarking (see Sec. 2.2).

Diamond Norm Distance and Channel Fidelity

The diamond norm of a linear superoperator $\mathcal{E} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$ is defined as,

$$\|\mathcal{E}\|_\diamond = \sup_{k\in\mathbb{N}} \|\mathcal{E} \otimes \mathcal{I}_k\|_1. \tag{A.39}$$

The 1-norm in Eq. (A.39) is the induced norm on linear superoperators by the 1-norm on the underlying space of linear operators. It is known that the supremum occurs for $k = m$ and so,

$$\|\mathcal{E}\|_\diamond = \|\mathcal{E} \otimes \mathcal{I}_m\|_1 = \max_{A\in L(\mathbb{C}^m\otimes\mathbb{C}^m):\|A\|_1\leq 1} \|\mathcal{E} \otimes \mathcal{I}_m(A)\|_1. \tag{A.40}$$

We note that the diamond norm is the dual of the cb-norm which is more common in the mathematical literature [112].

The diamond norm has the following operational meaning: Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be quantum channels from $L(\mathbb{C}^m)$ to $L(\mathbb{C}^n)$. Suppose we want to optimally distinguish between $\mathcal{E}_1$ and $\mathcal{E}_2$ using a single shot input state ρ under the assumption that the bit $a \in 1, 2$ is given to us uniformly at random from the uniform distribution on $\{1, 2\}$. More precisely, we want to minimize the error probability from a two-outcome POVM measurement on $\mathcal{E}_1(\rho)$ and $\mathcal{E}_2(\rho)$. In addition, we allow for possible entanglement with other quantum systems so that we actually want to minimize the error probability, p_E, from a two-outcome POVM measurement on $\mathcal{E}_1 \otimes \mathcal{I}_k(\rho)$ and $\mathcal{E}_2 \otimes \mathcal{I}_k(\rho)$. Thus for each k we want to find the state ρ which allows for optimal discrimination between the states $\mathcal{E}_1 \otimes \mathcal{I}_k(\rho)$ and $\mathcal{E}_2 \otimes \mathcal{I}_k(\rho)$, and then we want the supremum over all possible k. Assuming k and ρ are fixed, it is known that the minimum error probability in distinguishing between $\mathcal{E}_1 \otimes \mathcal{I}_k(\rho)$ and $\mathcal{E}_2 \otimes \mathcal{I}_k(\rho)$ is given by,

$$\frac{1}{2}\left(1 - \frac{1}{2}\|\mathcal{E}_1 \otimes \mathcal{I}_k(\rho) - \mathcal{E}_2 \otimes \mathcal{I}_k(\rho)\|_1\right). \tag{A.41}$$

If we want to find the minimum error probability p_E over all possible input states and over all k then we have,

$$p_E = \frac{1}{2}\left(1 - \frac{1}{2}\sup_{k\in\mathbb{N}}\max_{\rho\in D\left(L\left(\mathbb{C}^m\otimes\mathbb{C}^k\right)\right)}\|\mathcal{E}_1\otimes\mathcal{I}_k(\rho) - \mathcal{E}_2\otimes\mathcal{I}_k(\rho)\|_1\right). \tag{A.42}$$

From [142] since $\mathcal{E}_1$ and $\mathcal{E}_2$ are quantum operations, $\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond$ achieves its value at a quantum state, in fact a pure quantum state, in $L\left(\mathbb{C}^m\otimes\mathbb{C}^m\right)$. This implies from Eq. (A.39) that the diamond norm between the quantum operations can be written as,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \sup_{k\in\mathbb{N}}\max_{\rho\in D\left(L\left(\mathbb{C}^m\otimes\mathbb{C}^k\right)\right)}\|\mathcal{E}_1\otimes\mathcal{I}_k(\rho) - \mathcal{E}_2\otimes\mathcal{I}_k(\rho)\|_1. \tag{A.43}$$

Comparing the last two equations gives,

$$p_E = \frac{1}{2}\left(1 - \frac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond\right). \tag{A.44}$$

We can now explicitly see the operational significance of the diamond norm and a natural question to ask is whether allowing for entangled inputs does actually make a difference. The answer has been previously shown to be yes [78] and in the following proposition we show this fact using the specific case of depolarizing channels. Note that the semidefinite programming method used in Sec. 2.5.1 can also be used to show this result however we provide a proof that does not require knowledge of this method.

Proposition 6. *For depolarizing channels Λ_1 and Λ_2 on $L\left(\mathcal{H}\right)$ ($\dim\left(\mathcal{H}\right) = d$) with fidelity parameters p and q respectively,*

$$\|\Lambda_1 - \Lambda_2\|_1 = \frac{2(d-1)|p-q|}{d} \tag{A.45}$$

and,

$$\|\Lambda_1 - \Lambda_2\|_\diamond \geq \frac{2(d^2-1)|p-q|}{d^2}. \tag{A.46}$$

Proof. Let Λ_1 and Λ_2 be depolarizing channels on $L\left(\mathcal{H}\right)$:

$$\Lambda_1(\rho) = p\rho + (1-p)\mathrm{tr}(\rho)\frac{\mathbb{1}}{d} \tag{A.47}$$

and,

$$\Lambda_2(\rho) = q\rho + (1-q)\mathrm{tr}(\rho)\frac{\mathbb{1}}{d}. \tag{A.48}$$

For any pure state $|\psi\rangle\langle\psi| \in L(\mathcal{H})$ we have,

$$\begin{aligned} \|\Lambda_1(|\psi\rangle\langle\psi|) - \Lambda_2(|\psi\rangle\langle\psi|)\|_1 &= |p-q| \left\| |\psi\rangle\langle\psi| - \frac{\mathbb{1}}{d} \right\|_1 \\ &= 2|p-q| \left(\frac{d-1}{d}\right) \end{aligned} \tag{A.49}$$

and so,

$$\|\Lambda_1 - \Lambda_2\|_1 = 2|p-q| \left(\frac{d-1}{d}\right). \tag{A.50}$$

Now let us look at the diamond norm between Λ_1 and Λ_2,

$$\|\Lambda_1 - \Lambda_2\|_\diamond = \max_{|\psi\rangle \in \mathcal{H}\otimes\mathcal{H}} \|\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \Lambda_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1. \tag{A.51}$$

Let $|\psi\rangle \in \mathcal{H}\otimes\mathcal{H}$. It is straightforward to show that

$$\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|) = p|\psi\rangle\langle\psi| + (1-p)\frac{\mathbb{1}}{d} \otimes \mathrm{tr}_1(|\psi\rangle\langle\psi|) \tag{A.52}$$

where we recall $\mathrm{tr}_1(|\psi\rangle\langle\psi|)$ is the partial trace of $|\psi\rangle\langle\psi|$ over the first subsystem. Indeed we have for $|a\rangle|b\rangle\langle c|\langle d|$,

$$\begin{aligned} \Lambda_1 \otimes \mathcal{I}(|a\rangle|b\rangle\langle c|\langle d|) &= \Lambda_1(|a\rangle\langle c|) \otimes |b\rangle\langle d| \\ &= \left(p_1|a\rangle\langle c| + (1-p_1)\delta_{a,c}\frac{\mathbb{1}}{d}\right) \otimes |b\rangle\langle d| \\ &= p_1|a\rangle|b\rangle\langle c|\langle d| + (1-p_1)\frac{\mathbb{1}}{d} \otimes \mathrm{tr}_1(|a\rangle|b\rangle\langle c|\langle d|). \end{aligned} \tag{A.53}$$

So by linearity, $\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|) = p_1|\psi\rangle\langle\psi| + (1-p_1)\frac{\mathbb{1}}{d} \otimes \mathrm{tr}_1(|\psi\rangle\langle\psi|)$.

Let $\tau = \mathrm{tr}_1(|\psi\rangle\langle\psi|)$. Then,

$$\begin{aligned} \|\Lambda_1 - \Lambda_2\|_\diamond &\geq \left\| p|\psi\rangle\langle\psi| + (1-p)\frac{\mathbb{1}}{d} \otimes \tau - q|\psi\rangle\langle\psi| - (1-q)\frac{\mathbb{1}}{d} \otimes \tau \right\|_1 \\ &= |p-q| \left\| |\psi\rangle\langle\psi| - \frac{\mathbb{1}}{d} \otimes \tau \right\|_1 \end{aligned} \tag{A.54}$$

and note that in no case are $|\psi\rangle\langle\psi|$ and $\frac{\mathbb{1}}{d} \otimes \tau$ orthogonal since $\mathrm{tr}\left(|\psi\rangle\langle\psi| \left(\frac{\mathbb{1}}{d} \otimes \tau\right)\right) = \frac{\mathrm{tr}(\tau^2)}{d}$. We want to maximize the value of $\||\psi\rangle\langle\psi| - \frac{\mathbb{1}}{d} \otimes \tau\|_1$. Intuitively, the maximum occurs when $\tau = \frac{\mathbb{1}}{d}$ (ie. $|\psi\rangle$ is a maximally entangled state which we choose to be $|\psi_0\rangle$) so that

$$\begin{aligned} \left\| |\psi\rangle\langle\psi| - \frac{\mathbb{1}}{d} \otimes \tau \right\|_1 &= \left\| |\psi_0\rangle\langle\psi_0| - \frac{\mathbb{1}}{d} \otimes \frac{\mathbb{1}}{d} \right\|_1 \\ &= \frac{2\,(d^2-1)}{d^2}. \end{aligned} \tag{A.55}$$

Hence,

$$\begin{aligned} \|\Lambda_1 - \Lambda_2\|_\diamond &\geq |p-q| \left\| |\psi_0\rangle\langle\psi_0| - \frac{\mathbb{1}}{d} \otimes \frac{\mathbb{1}}{d} \right\|_1 \\ &= \frac{2(d^2-1)|p-q|}{d^2}. \end{aligned} \tag{A.56}$$

Thus,

$$\|\Lambda_1 - \Lambda_2\|_\diamond \geq \frac{2(d^2-1)|p-q|}{d^2} > \frac{2(d-1)|p-q|}{d} = \|\Lambda_1 - \Lambda_2\|_1. \tag{A.57}$$

Thus ancilla systems make a difference.

□

We now define the channel fidelity between quantum operations and then discuss some relationships between the diamond norm distance and channel fidelity.

Definition 7. *Channel Fidelity*

The channel fidelity between two quantum operations $\mathcal{E}_1$ and $\mathcal{E}_2$ in $\mathcal{S}(\mathcal{H}_2, \mathcal{H}_2)$ is the real-valued function on quantum states given by

$$\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}(\rho) = \left(tr\sqrt{\sqrt{\mathcal{E}_1(\rho)}\mathcal{E}_2(\rho)\sqrt{\mathcal{E}_1(\rho)}} \right)^2$$

where ρ is an arbitrary mixed quantum state in $\mathcal{D}(\mathcal{H}_1)$.

Note that the channel fidelity is a state-dependent measure of the distance between two quantum channels, which is a simple sign that it is not a metric. We have the following result relating the diamond norm distance and channel fidelity.

Theorem 4. *For any quantum operations Λ_1 and Λ_2 in $\mathcal{S}(\mathcal{H}_1, \mathcal{H}_2)$,*

$$min_{\rho\in\mathcal{D}(\mathcal{H}_1)} F(\Lambda_1(\rho), \Lambda_2(\rho)) \geq 1 - \|\Lambda_1 - \Lambda_2\|_\diamond \tag{A.58}$$

where $min_{\rho\in\mathcal{D}(\mathcal{H}_1)} F(\Lambda_1(\rho), \Lambda_2(\rho))$ is the minimum channel fidelity between Λ_1 and Λ_2. In the case that one of Λ_1 or Λ_2 is a unitary operation, this is the familiar minimum gate fidelity (see Sec A.4.2).

Proof. We have that,

$$\|\Lambda_1 - \Lambda_2\|_\diamond = \max_{|\psi\rangle\in\mathcal{H}_1\otimes\mathcal{H}_1} \|\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \Lambda_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1. \tag{A.59}$$

By the Fuchs-Van de Graaf inequalities [52],

$$\|\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \Lambda_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1 \geq 1 - F(\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|), \Lambda_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)) \tag{A.60}$$

so

$$\begin{aligned} \|\Lambda_1 - \Lambda_2\|_\diamond &\geq \max_{|\psi\rangle\in\mathcal{H}_1\otimes\mathcal{H}_1} \left[1 - F(\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|), \Lambda_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|))\right] \\ &= 1 - \min_{|\psi\rangle\in\mathcal{H}_1\otimes\mathcal{H}_1} F(\Lambda_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|), \Lambda_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)). \end{aligned} \tag{A.61}$$

Now, we prove the following proposition:

Proposition 7.

$$min_{|\psi\rangle\in\mathcal{H}_1\otimes\mathcal{H}_1} F_{\Lambda_1\otimes\mathcal{I},\Lambda_2\otimes\mathcal{I}}(|\psi\rangle\langle\psi|) \leq min_{|\phi\rangle\in\mathcal{H}_1} F_{\Lambda_1,\Lambda_2}(|\phi\rangle\langle\phi|) \tag{A.62}$$

that is,

$$min_{|\psi\rangle\in\mathcal{H}_1\otimes\mathcal{H}_1} F(\Lambda_1\otimes\mathcal{I}(|\psi\rangle\langle\psi|), \Lambda_2\otimes\mathcal{I}(|\psi\rangle\langle\psi|)) \leq min_{|\phi\rangle\in\mathcal{H}_1} F(\Lambda_1(|\phi\rangle\langle\phi|), \Lambda_2(|\phi\rangle\langle\phi|)) \tag{A.63}$$

Proof. We have,

$$\begin{aligned} \min_{|\psi\rangle\in\mathcal{H}_1\otimes\mathcal{H}_1} F_{\Lambda_1\otimes\mathcal{I},\Lambda_2\otimes\mathcal{I}}(|\psi\rangle\langle\psi| &\leq \min_{|\phi\rangle\in\mathcal{H}_1} F(\Lambda_1\otimes\mathcal{I}(|\phi\rangle\langle\phi|\otimes|\phi\rangle\langle\phi|), \Lambda_2\otimes\mathcal{I}(|\phi\rangle\langle\phi|\otimes|\phi\rangle\langle\phi|)) \\ &= \min_{|\phi\rangle\in\mathcal{H}_1} F(\Lambda_1(|\phi\rangle\langle\phi|)\otimes|\phi\rangle\langle\phi|), \Lambda_2(|\phi\rangle\langle\phi|)\otimes|\phi\rangle\langle\phi|)). \end{aligned} \tag{A.64}$$

By definition this is equal to

$$\min_{|\phi\rangle\in\mathcal{H}_1} \left(\mathrm{tr}\sqrt{\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)\otimes|\phi\rangle\langle\phi|}\,(\Lambda_2(|\phi\rangle\langle\phi|)\otimes|\phi\rangle\langle\phi|)\,\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)\otimes|\phi\rangle\langle\phi|}} \right)^2 \tag{A.65}$$

which is,

$$\min_{|\phi\rangle\in\mathcal{H}_1} \left(\mathrm{tr}\sqrt{\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)}\otimes|\phi\rangle\langle\phi|\,(\Lambda(|\phi\rangle\langle\phi|)\otimes|\phi\rangle\langle\phi|)\,\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)}\otimes|\phi\rangle\langle\phi|} \right)^2. \tag{A.66}$$

However the above is equal to

$$\min_{|\phi\rangle\in\mathcal{H}_1} \left(\mathrm{tr}\sqrt{\left(\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)}\,(\Lambda_2(|\phi\rangle\langle\phi|))\,\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)}\right)\otimes|\phi\rangle\langle\phi|} \right)^2 \tag{A.67}$$

which is just,

$$\min_{|\phi\rangle\in\mathcal{H}_1} F_{\Lambda_1,\Lambda_2}(|\phi\rangle\langle\phi|) = \min_{|\phi\rangle\in\mathcal{H}_1} \left(\mathrm{tr}\sqrt{\left(\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)}\,(\Lambda_2(|\phi\rangle\langle\phi|))\,\sqrt{\Lambda_1(|\phi\rangle\langle\phi|)}\right)} \right)^2. \tag{A.68}$$

which proves the proposition.

□

So,

$$\|\Lambda_1 - \Lambda_2\|_\diamond \geq 1 - \min_{|\phi\rangle\in\mathcal{H}_1} F(\Lambda_1(|\phi\rangle\langle\phi|), \Lambda_2(|\phi\rangle\langle\phi|)). \tag{A.69}$$

Now let $\rho = \sum_i p_i |\alpha_i\rangle\langle\alpha_i| \in \mathcal{D}(\mathcal{H}_1)$. By joint concavity of the fidelity,

$$\begin{aligned} F(\Lambda_1(\rho), \Lambda_2(\rho)) &= F\left(\Lambda_1\left(\sum_i p_i|\alpha_i\rangle\langle\alpha_i|\right), \Lambda_2\left(\sum_i p_i|\alpha_i\rangle\langle\alpha_i|\right)\right) \\ &= F\left(\sum_i p_i \Lambda_1\left(|\alpha_i\rangle\langle\alpha_i|\right), \sum_i p_i \Lambda_2\left(|\alpha_i\rangle\langle\alpha_i|\right)\right) \\ &\geq \sum_i p_i F\left(\Lambda_1\left(|\alpha_i\rangle\langle\alpha_i|\right), \Lambda_2\left(|\alpha_i\rangle\langle\alpha_i|\right)\right) \\ &\geq \sum_i p_i \min_{|\alpha_j\rangle} F(\Lambda_1(|\alpha_j\rangle\langle\alpha_j|), \Lambda_2(|\alpha_j\rangle\langle\alpha_j|)) \\ &= \min_{|\alpha_j\rangle} F(\Lambda_1(|\alpha_j\rangle\langle\alpha_j|), \Lambda_2(|\alpha_j\rangle\langle\alpha_j|)). \end{aligned} \tag{A.70}$$

Thus we get,

$$\min_{\rho\in\mathcal{D}(\mathcal{H}_1)} F(\Lambda_1(\rho), \Lambda_2(\rho)) = \min_{|\phi\rangle\in\mathcal{H}_1} F(\Lambda_1(|\phi\rangle\langle\phi|), \Lambda_2(|\phi\rangle\langle\phi|)) \tag{A.71}$$

and so,

$$\min_{\rho\in\mathcal{D}(\mathcal{H}_1)} F(\Lambda_1(\rho),\Lambda_2(\rho)) \geq 1 - \|\Lambda_1 - \Lambda_2\|_\diamond. \tag{A.72}$$

□

More Distance Measures: Schatten Norms and the Gate Fidelity

There are many other methods for quantifying the distance between linear superoperators. One particularly useful method is given by the set of superoperator norms that are induced by the Schatten p-norms on the underlying operator space [142]. In particular, we focus on the $\|\ \|_{1\to 1}^H$ norm which is defined for linear superoperator $\mathcal{R} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$ as,

$$\|\mathcal{R}\|_{1\to 1}^H = \max_{A:A=A^\dagger,\|A\|_1\leq 1} \|\mathcal{R}(A)\|_1 \tag{A.73}$$

where $A \in L(\mathbb{C}^m)$. One can see that $\|\ \|_{1\to 1}^H$ is just $\|\ \|_1$ (which is also denoted $\|\ \|_{1\to 1}$) restricted to Hermitian inputs. This norm is less common in quantum information due to its lack of operational meaning, however it is a weaker measure of distance than the diamond norm since for any linear superoperator $\mathcal{R} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$, $\|\mathcal{R}\|_{1\to 1}^H \leq \|\mathcal{R}\|_\diamond$. This will be of particular use when we consider neglecting higher order effects in the benchmarking scheme (see Sec. 2.2).

In the case of a unitary operation $\mathcal{U}$, quantum operation $\mathcal{E}$, and restricting input states to $\mathbb{CP}^{d-1}$, the channel fidelity introduced in the previous section is called the gate fidelity.

Definition 8. *Gate Fidelity*

The gate fidelity between quantum operation $\mathcal{E}$ and unitary $\mathcal{U}$ is the real-valued function on pure quantum states given by

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(\phi) = tr\left(\mathcal{U}(|\phi\rangle\langle\phi|)\mathcal{E}(|\phi\rangle\langle\phi|)\right), \tag{A.74}$$

and defining $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$ gives,

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(\phi) = \mathcal{F}_{\Lambda,\mathcal{I}}(\phi) = tr\left(|\phi\rangle\langle\phi|\Lambda(|\phi\rangle\langle\phi|)\right). \tag{A.75}$$

The channel Λ can be thought of as representing how much $\mathcal{E}$ deviates from $\mathcal{U}$ in that if $\mathcal{E} = \mathcal{U}$ then $\Lambda = \mathcal{I}$. The gate fidelity has many nice mathematical properties including a simple

expression for the average over pure states, expressions for the variance in terms of various representations of Λ and a concentration of measure phenomenon for large systems [107, 98, 97].

The average gate fidelity is obtained by integrating $\mathcal{F}_{\mathcal{E},\mathcal{U}}$ over $\mathbb{CP}^{d-1}$ using the Fubini-Study measure μ_{FS} [12],

$$\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}} = \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} = \int_{\mathbb{CP}^{d-1}} \mathrm{tr}\left(|\phi\rangle\langle\phi|\Lambda(|\phi\rangle\langle\phi|)\right) d\mu_{FS}(\phi). \tag{A.76}$$

Note that if $\mathcal{E}$ is depolarizing with $\mathcal{E}(\rho) = p\rho + (1-p)\frac{\mathbb{1}}{d}$ and $\mathcal{U} = \mathcal{I}$ then for every pure state $|\phi\rangle\langle\phi|$,

$$\mathcal{F}_{\mathcal{E},\mathcal{U}}(|\phi\rangle\langle\phi|) = p + \frac{1-p}{d}. \tag{A.77}$$

Taking the minimum of $\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}$ over all mixed states ρ produces a quantity $\mathcal{F}^{\min}_{\mathcal{E}_1,\mathcal{E}_2}$ commonly called the minimum channel fidelity,

$$\mathcal{F}^{\min}_{\mathcal{E}_1,\mathcal{E}_2} = \min_\rho \mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}(\rho).$$

Note that by concavity of the fidelity, the minimum channel fidelity occurs at a pure state [108]. In the case of the gate fidelity, the minimum is called the minimum gate fidelity.

In certain cases we will be concerned with how close $\mathcal{E}_1$ and $\mathcal{E}_2$ are in terms of the difference between the average fidelity of each channel. To this end we define,

$$\Delta F(\mathcal{E}_1, \mathcal{E}_2) := \left|\overline{\mathcal{F}_{\mathcal{E}_1,\mathcal{I}}} - \overline{\mathcal{F}_{\mathcal{E}_2,\mathcal{I}}}\right|. \tag{A.78}$$

We note the following relationships between some of the distance measures defined above. First, for $\mathcal{E}_1$, $\mathcal{E}_2 \in \mathcal{S}(\mathcal{H})$ the following inequalities hold,

$$\Delta F(\mathcal{E}_1, \mathcal{E}_2) \leq \|\mathcal{E}_1 - \mathcal{E}_2\|^H_{1\to 1} \leq \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond \tag{A.79}$$

where we recall the definition of $\|\ \|^H_{1\to 1}$ in Eq. (A.73). The second inequality is clear since,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to 1}^H \leq \|\mathcal{E}_1 - \mathcal{E}_2\|_1 \leq \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond. \tag{A.80}$$

Now for the first inequality we have that,

$$\begin{aligned}
\Delta F(\mathcal{E}_1, \mathcal{E}_2) &\leq \max_{|\phi\rangle} \left|\mathrm{tr}\left(\left(\mathcal{E}_1 - \mathcal{E}_2\right)\left(|\phi\rangle\langle\phi|\right)|\phi\rangle\langle\phi|\right)\right| \\
&\leq \max_{|\phi\rangle,|\beta\rangle} \left|\mathrm{tr}\left(\left(\mathcal{E}_1 - \mathcal{E}_2\right)\left(|\phi\rangle\langle\phi|\right)|\beta\rangle\langle\beta|\right)\right| \\
&= \max_{|\phi\rangle} \left(\max\{|\sigma_i| \ : \sigma_i \text{ is an eigenvalue of} \left(\Lambda - \Lambda_i\right)\left(|\phi\rangle\langle\phi|\right)\}\right) \\
&= \max_{|\phi\rangle} \rho\left(\left(\mathcal{E}_1 - \mathcal{E}_2\right)\left(|\phi\rangle\langle\phi|\right)\right) \\
&= \max_{|\phi\rangle} \left\|\left(\mathcal{E}_1 - \mathcal{E}_2\right)\left(|\phi\rangle\langle\phi|\right)\right\|_\infty \\
&= \max_{A:A=A^\dagger,\|A\|_1\leq 1} \left\|\left(\mathcal{E}_1 - \mathcal{E}_2\right)(A)\right\|_\infty \\
&= \|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to\infty}^H
\end{aligned} \tag{A.81}$$

where since $\mathcal{E}_1$ and $\mathcal{E}_2$ are completely positive, $\mathcal{E}_1 - \mathcal{E}_2$ is Hermiticity-preserving and "$\rho(\,\cdot\,)$" represents the spectral radius of a linear operator. Hence since $\|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to\infty}^H \leq \|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to 1}^H$, the inequalities in Eq. (A.79) hold.

It is also of interest to obtain bounds on the diamond norm distance and the average fidelity. Ref. [10] gives that for any two quantum operations $\mathcal{E}_1$ and $\mathcal{E}_2 \in \mathcal{S}(\mathcal{H}_1, \mathcal{H}_2)$,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond \leq 2d\|J(\mathcal{E}_1) - J(\mathcal{E}_2)\|_1. \tag{A.82}$$

Let us set $d_1 = d_2 = d$, $\mathcal{E}_2 = \mathcal{I}$, and $\mathcal{E}_1 = \mathcal{E}$. It is also shown that

$$\begin{aligned}
\frac{1}{2}\|J(\mathcal{E}) - J(\mathcal{I})\|_1 &= \frac{1}{2}\|J(\mathcal{E}) - |\psi_0\rangle\langle\psi_0|\|_1 \\
&\leq \sqrt{1 - \langle\psi_0|J(\mathcal{E})|\psi_0\rangle}
\end{aligned} \tag{A.83}$$

where $\langle\psi_0|J(\mathcal{E})|\psi_0\rangle$ is the entanglement fidelity of $\mathcal{E}$ (denoted hereafter by $\mathcal{F}_{\mathrm{ent}}(\mathcal{E})$) which is related to $\overline{\mathcal{F}_{\mathcal{E},\mathcal{I}}}$ by [108]

$$\mathcal{F}_{\mathrm{ent}}(\mathcal{E}) = \frac{(d+1)\overline{\mathcal{F}_{\mathcal{E},\mathcal{I}}} - 1}{d}. \tag{A.84}$$

Thus,

$$\frac{1}{4d}\|\mathcal{E}-\mathcal{I}\|_\diamond \;\leq\; \sqrt{1-\mathcal{F}_{\text{ent}}(\mathcal{E})} \tag{A.85}$$

which implies after some algebra,

$$\|\mathcal{E}-\mathcal{I}\|_\diamond \;\leq\; 4\sqrt{d(d+1)(1-\overline{\mathcal{F}_{\mathcal{E},\mathcal{I}}})}. \tag{A.86}$$

Lastly, six properties that a useful measure of distance, β, should satisfy are discussed in [56] and listed here for reference,

1. *Metric*: β should be a metric.

2. *Easy to calculate*: There should be a straightforward method for evaluating β.

3. *Easy to measure*: There should be a clear and achievable experimental protocol for determining β.

4. *Physical interpretation*: β should have a well-motivated physical interpretation.

5. *Stability*: β should be stable under tensoring with the identity operation, ie. if $\mathcal{Q}$ and $\mathcal{R}$ are quantum operations, $\beta\left(\mathcal{Q}\otimes\mathcal{I},\mathcal{R}\otimes\mathcal{I}\right)=\beta\left(\mathcal{Q},\mathcal{R}\right)$.

6. *Chaining*: For a process composed of many smaller steps, the total error will be less than the sum of the errors in the individual steps, ie. for channels $\mathcal{Q}_1$, $\mathcal{Q}_2$, $\mathcal{R}_1$ and $\mathcal{R}_2$, $\beta(\mathcal{Q}_2\circ\mathcal{Q}_1,\mathcal{R}_2\circ\mathcal{R}_1)\leq\beta(\mathcal{Q}_2,\mathcal{R}_2)+\beta(\mathcal{Q}_1,\mathcal{R}_1)$.

$\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$ and$\mathcal{F}^{\min}_{\mathcal{E},\mathcal{U}}$ are both candidates to be a good measure of distance. $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$ is shown in [56] to satisfy properties 2, 3 and 4 but fails to satisfy the rest. $\mathcal{F}^{\min}_{\mathcal{E},\mathcal{U}}$ on the other hand satisfies all of the properties except for 2 and 3. It should be noted that if process tomography can be performed then $\mathcal{F}^{\min}_{\mathcal{E},\mathcal{U}}$ can be calculated numerically using convex optimization techniques.

Appendix B

Quantum Information Theory

The fundamental unit of information in quantum information theory is a quantum bit or "qubit", analogous to the "bit" in classical information theory. Physically, a qubit may by thought of as a two-dimensional quantum mechanical system. Hence it is mathematically represented by the set of trace 1 positive operators acting on a 2-dimensional complex Hilbert space. A standard physical instance of a qubit is given by photon polarization, where three sets of bases for the system are the horizontal-vertical (H/V) basis, plus-minus basis (+/-) and the right-left circular polarization (R/L) basis.

From Sec. A.1.1, the state space for two qubits, each with basis $\{|0\rangle, |1\rangle\}$, has basis $\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$ which will be written more compactly as $\{|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle\}$. In general an n qubit system has 2^n basis vectors. The orthonormal basis for an n qubit Hilbert space formed from tensor products of $|0\rangle$ and $|1\rangle$ is called the computational basis. More generally, we write $|x\rangle$ to mean $|b_n b_{n-1} \dots b_0\rangle$ where b_i are the binary digits of the number x.

The maximally entangled state $|\psi_0\rangle$ discussed previously is given by $\frac{|00\rangle+|11\rangle}{2}$ for a two-qubit system. There exists an orthonormal basis $\mathcal{B}$ for a two-qubit system that consists only of maximally entangled states:

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle), \frac{1}{\sqrt{2}} (|00\rangle - |11\rangle), \frac{1}{\sqrt{2}} (|01\rangle + |10\rangle), \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle) \right\}. \quad \text{(B.1)}$$

$\mathcal{B}$ is commonly called the "Bell" basis and will be of significant importance to us in this thesis.

B.1 Quantum Gates

As previously mentioned, the dynamics of a quantum system, when not interacting with an environment or being measured, is described by a unitary transformation. One important consequence of the fact that quantum transformations are unitary is that they are reversible.

B.1.1 Pauli Operators

The Pauli operators are extremely important in quantum information processing and are the single-qubit unitary transformations written in the basis $\{|0\rangle, |1\rangle\}$ as,

$$\begin{array}{lrcll} \mathbb{1}: & |0\rangle & \rightarrow & |0\rangle & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ & |1\rangle & \rightarrow & |1\rangle & \\ X: & |0\rangle & \rightarrow & |1\rangle & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ & |1\rangle & \rightarrow & |0\rangle & \\ Y: & |0\rangle & \rightarrow & i\,|1\rangle & \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ & |1\rangle & \rightarrow & -i\,|0\rangle & \\ Z: & |0\rangle & \rightarrow & |0\rangle & \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \\ & |1\rangle & \rightarrow & -|1\rangle & \end{array}$$

The set $\{\mathbb{1}, X, Y, Z\}$ forms a traceless (except for $\mathbb{1}$), unitary, Hermitian and orthogonal basis for $L(\mathbb{C}^2)$ (orthonormality is attained when each element is scaled by $\frac{1}{2}$). $\mathbb{1}, X, Y$ and Z are sometimes denoted either as P_0, P_1, P_2, and P_3 or σ_0, σ_1, σ_2 and σ_3, and they satisfy the commutation/anti-commutation relations

$$[P_l, P_m] = 2i \sum_{n=1}^{3} \epsilon_{lmn} P_n, \tag{B.2}$$

$$\{P_l, P_m\} = 2\delta_{l,m} \mathbb{1}. \tag{B.3}$$

If phases in $\{1, -1, i, -i\}$ are allowed, the set of Pauli operators forms a group under multiplication which we denote by $\mathcal{P}_1$. More generally, for any n, the set of n-fold tensor

products of Pauli's forms a traceless, unitary, Hermitian orthogonal basis for $L\left((\mathbb{C}^2)^{\otimes n}\right)$. As well, if phases in $\{1, -1, i, -i\}$ are allowed, this set forms a group under multiplication which we denote by $\mathcal{P}_n$.

The Bloch sphere [16] is a useful representation of quantum states. For a single-qubit state ρ we can write

$$\rho = \frac{1}{2}\left(\sigma_0 + \sum_{j=1}^{3} r_j \sigma_j\right) \tag{B.4}$$

where the 3-vector $\vec{r}$ lies in the unit sphere of $\mathbb{R}^3$. This vector is called the Bloch vector (representation) of ρ. For single qubit states, every point in the unit sphere is associated to a unique quantum state and the boundary (shell) of the unit sphere corresponds exactly to the set of pure states.

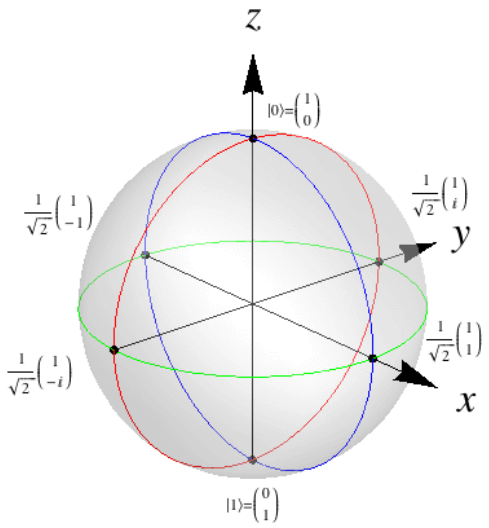

Figure B.1: Bloch sphere representation of single-qubit states.

The generalization of the Bloch vector to multi-qubit systems is straightforward. Unfortunately, the pictorial representation of states being associated to every point on the sphere breaks down due to the exotic geometry of multi-qubit state spaces [12].

Quantum channels also take a simple form in the Bloch sphere representation [19, 53, 123]. Any quantum channel Λ can be uniquely represented by a real matrix M_Λ and fixed

vector $\vec{t}$ such that

$$\vec{r} \to M_\Lambda \vec{r} + \vec{t}. \tag{B.5}$$

Since $\vec{t}$ is fixed, this is an affine transformation and $\vec{t}$ characterizes the non-unitality of the channel. The Bloch representation preserves many of the intuitive features of quantum operations, for instance a unitary operation $\mathcal{U}$ is represented by an orthogonal (rotation) matrix $M_\mathcal{U}$ and Pauli channels are represented by diagonal matrices.

Generalized bases

For qubits ($d = 2$), the Pauli operators are an exceptionally convenient basis for $L(\mathcal{H})$. The corresponding basis of bipartite states (see Sec. A.2) is the Bell basis. In dimensions not equal to 2^n for some n it is generally not possible to pick a basis with *all* the nice properties of the Pauli operators, but we can generalize most of them.

We will make extensive use of the existence of a Hermitian, orthogonal basis of matrices $\{P_a\}$ satisfying the following conditions:

$$\begin{aligned} \operatorname{tr}(P_a P_b) &= d\delta_{a,b} \\ P_a^\dagger &= P_a \\ P_0 &= \mathbb{1}. \end{aligned} \tag{B.6}$$

In any dimension, the generalized Gell-Mann operators [54] satisfy these conditions. Note that $P_0 = \mathbb{1}$ and therefore every other P_k is traceless. The corresponding basis of bipartite states $\{(P_a \otimes \mathbb{1}) |\psi_0\rangle\}$ is orthonormal. We will refer to this basis as the "generalized Bell basis," though it does not by any means generalize *all* the properties of the Bell states.

We will also make extensive use of the bipartite projector

$$\chi_0 = |\psi_0\rangle\langle\psi_0|.$$

It is proportional to the Choi representation of the identity channel $\mathcal{E} = \mathbb{1}$, which motivates the notation χ_0. Moreover, it enables us to write expressions we derive using the basis $\{P_a\}$ in terms of quantities that are defined independently of any basis – e.g. $\chi_{0,0}$ mentioned above can be written as

$$\chi_{0,0} = \langle\psi_0| \chi |\psi_0\rangle = \operatorname{tr}[\chi\chi_0].$$

B.1.2 More Gates

Another useful single qubit transformation is the Hadamard Transformation defined by

$$\begin{array}{rccl} H: & |0\rangle & \rightarrow & \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = |+\rangle \\ & |1\rangle & \rightarrow & \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = |-\rangle. \end{array}$$

The transformation H has a number of important applications. For instance, when applied to $|0\rangle$, H creates a superposition state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Applied to n qubits individually, H generates a superposition of all 2^n possible states, which can be viewed as the binary representation of the numbers from 0 to $2^n - 1$,

$$\begin{aligned} H \otimes H \otimes \cdots \otimes H \, |00\ldots0\rangle &= \frac{1}{\sqrt{2^n}} \left((|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes \cdots \otimes (|0\rangle + |1\rangle) \right) \\ &= \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle \, . \end{aligned} \tag{B.7}$$

The family $\{R_k\}$ of unitary transformations on a single qubit defined in the standard basis by

$$R_k = \begin{bmatrix} 1 & 0 \\ 0 & e^{-\frac{2\pi i}{2^k}} \end{bmatrix}$$

is also of great use in quantum information. The particular case of $k = 1$ is called the phase gate and is denoted S.

The controlled-NOT gate, $CNOT$, operates on two qubits as follows: it flips the second qubit if the first qubit is $|1\rangle$ and leaves the second qubit unchanged when the first is $|0\rangle$. As noted, the vectors $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ form an orthonormal basis for the set of pure states for a two-qubit system. Hence the $CNOT$ transformation has representation in this basis given by

$$\begin{array}{rccl} CNOT: & |00\rangle & \rightarrow & |00\rangle \\ & |01\rangle & \rightarrow & |01\rangle \\ & |10\rangle & \rightarrow & |11\rangle \\ & |11\rangle & \rightarrow & |10\rangle \end{array} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The transformation $CNOT$ is unitary and cannot be decomposed into a tensor product of

two single qubit transformations.

It is useful to have graphical representations of quantum state transformations, especially when several transformations are combined in sequence. This representation is given by a quantum circuit, which is read left to right in time. The number of horizontal levels in the circuit corresponds to the number of qubits involved in the computation. The following is an example of a quantum circuit.

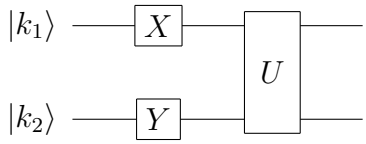

Figure B.2: Example of a quantum circuit

There are two qubits, the first in the state $|k_1\rangle$ and the second in the state $|k_2\rangle$. The first qubit undergoes the unitary transformation X and the second is transformed by Y. The entire 2 qubit system then undergoes the unspecified unitary transformation U.

$CNOT$ is typically represented by a circuit of the form

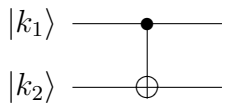

Figure B.3: $CNOT$ gate

The filled circle indicates the control qubit, and the $\oplus$ indicates the conditional negation of the target qubit.

B.2 The Clifford Group, Universality and Quantum Algorithms

The Clifford group on n qubits, denoted Clif_n, is defined as the normalizer of the Pauli group $\mathcal{P}_n$ and plays an important role in many areas of quantum information such as universality [20], stabilizer code theory [57] and noise estimation [40]. A set of gates is said to be universal for quantum computation if any unitary operator can be approximated to

arbitrary accuracy using only gates in this set. A universal set of gates on n qubits cannot be generated from Clif_n alone, however the addition of a single element not in Clif_n, such as the $\frac{\pi}{8}$ gate,

$$\frac{\pi}{8} = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{4}} \end{bmatrix}$$

produces a universal set for the full unitary group $U(2^n)$ [20]. Using the fact that Clif_n is generated by H and S on each qubit, coupled with $CNOT$ gates on all pairs of qubits, a discrete universal set for $U(2^n)$ exists and is given by H, S and $\frac{\pi}{8}$ on each qubit coupled with $CNOT$ on all pairs of qubits. It can be shown that the $\frac{\pi}{8}$ gate can be implemented via a teleportation scheme by preparing an ancilla in a specific initial "magic" state [21], performing a measurement in the computational basis and using elements only from Clif_n [145]. One can therefore replace the requirement of adding $\frac{\pi}{8}$ to Clif_n for generating $U(2^n)$ with the ability to prepare an ancilla in a magic state and perform measurements in the computational basis.

Since the Clifford group is not universal, Clifford computation (ie. quantum computation using only Clifford elements and measurements of observables in the Pauli group) is not as powerful as quantum computation. An interesting question with important applications is where this class of computation lies between classical and quantum computation. If Clifford computation were more powerful than classical computation then conceivably one could solve problems on a Clifford quantum computer that may not be tractable on a classical computer. Interestingly though, the Gottesman-Knill theorem [57] shows that Clifford circuits can be efficiently simulated on a classical computer, hence Clifford computation is no more powerful than classical computation.

One of the main advantages of representing information through quantum systems is that certain computational problems with no known efficient classical solution are efficiently solvable using a quantum information processor. Many of the quantum algorithms that solve these problems rely on a transformation called the quantum Fourier transform [75, 108]. One of the most prominent computational problem with wide-ranging applications is factoring integers (RSA). Shor's quantum algorithm for solving the factoring problem [131] is essentially a specific case of a more general problem called the hidden subgroup problem [76]. Algorithms for solving various instances of the hidden subgroup problem, such as the case of Abelian groups, rely heavily on the representation theory of finite groups. While it is useful to know that certain computational tasks are easy on a quantum information processor, implementing such computations is extremely hard. This difficulty is due to the extreme sensitivity of quantum systems to their environment. The area of research that

deals with reliably preserving information when a quantum system interacts with some environment is called quantum error-correction and will be examined next.

B.3 Quantum Error Correction

Quantum error correction (QEC) is a subfield of quantum information theory that deals with how to preserve quantum information when it is sent through a channel. Representing information through quantum states suffers from the drawback that a quantum system is extremely sensitive to interactions with an environment. These interactions create correlations between the system of interest and the environment which results in the environment carrying away information about the system. Thus, information initially encoded in the quantum system may be lost through such interactions. Sending information from one party to another requires that the received state of the system closely resembles the initial information. Hence we must find ways to minimize the interaction of an environment with the encoded information.

There are two types of error correction, passive and active. In passive error correction once the initial state has been encoded it only interacts with the quantum channel. Thus, the main part of the error correction procedure lies in the encoding and decoding of the quantum information. Active error correction pertains to actively manipulating the state while, or after, it interacts with the channel in order to preserve the encoded information. QEC through noiseless subsystems is a type of passive error correction where certain subsystems of the state space are located as being "unaffected" by the quantum channel. An example of active error correction is given by dynamical decoupling methods [141] which are of great utility in many implementations such as NMR.

A method for quantum error correction is given in [86] that unifies the previously known methods of error correction under one framework. This framework is called operator quantum error correction and applies to both unital and non-unital quantum channels. Some of the previous methods of error correction are the standard model, decoherence free subspaces (which includes stabilizer code theory [57]) and noiseless subsystems. We briefly discuss these methods for error correction and then present the unified method.

B.3.1 The Standard Model

The standard model may be described by a triple $(\mathcal{R}, \mathcal{E}, \mathcal{C})$, where the code $\mathcal{C}$ is a subspace of the Hilbert space $\mathcal{H}$, $\mathcal{E}$ is a quantum channel, and $\mathcal{R}$ is a recovery channel. Denote

the projection onto $\mathcal{C}$ by $P_{\mathcal{C}}$. The triple must satisfy the following for all bounded linear operators $\rho = P_{\mathcal{C}}\rho P_{\mathcal{C}}$ (ie. all ρ which are reduced by $P_{\mathcal{C}}$ and whose support lies in $\mathcal{C}$),

$$(\mathcal{R} \circ \mathcal{E})(\rho) = \rho. \tag{B.8}$$

When there exists an $\mathcal{R}$ for given $\mathcal{E}$ and $\mathcal{C}$, the code $\mathcal{C}$ is said to correct $\mathcal{E}$. In the case $\mathcal{R} = \mathcal{I}$, the triple is called a decoherence free subspace. Let $\{E_a\}$ be a set of Kraus operators for $\mathcal{E}$. Then the existence of $\mathcal{R}$ for $\mathcal{E}$ and $\mathcal{C}$ is equivalent to

$$P_{\mathcal{C}} E_a^{\dagger} E_b P_{\mathcal{C}} = \mu_{ab} P_{\mathcal{C}} \tag{B.9}$$

for all a, b in the index set for the Kraus representation where the matrix μ_{ab} is positive semi-definite with trace equal to 1 [108]. Since different Kraus representations for a particular CP map are related by a unitary matrix, the form of the above condition is clearly independent of the Kraus representation used.

By the unitary freedom in the Kraus operators and the fact that μ_{ab} is positive semi-definite, there exists a set of Kraus operators for the channel $\mathcal{E}$ such that μ_{ab} is diagonal. For this particular set of Kraus operators, labeled $\{G_a\}$, it is clear that the code subspace $\mathcal{C}$ is mapped to orthogonal subspaces by the G_a. So, $\mathcal{C}$ is correctable for $\mathcal{E}$ if and only if there exists a Kraus representation $\{G_a\}$ such that:

1. $\forall\ |\psi\rangle \in H$ and $a \neq b$, $G_a P\ |\psi\rangle$ is either equal or orthogonal to $G_b P\ |\psi\rangle$
2. The inner product structure on $\mathcal{C}$ is preserved by the G_a.

Thus $\mathcal{C}$ is mapped to orthogonal undeformed copies of $\mathcal{C}$ in H. This is a useful property of the $\{G_a\}$ because the recovery operation is then easily described by a measurement in a basis determined by the orthogonal copies followed by a unitary operation [108]. It is important to note that there may exist a set of Kraus operators for $\mathcal{E}$ such that the action of at least two of the Kraus operators on $\mathcal{C}$ is the same and $\mathcal{C}$ is still a correctable subspace for $\mathcal{E}$. This occurs when μ does not have maximal rank and this phenomenon is called degeneracy with $\mathcal{C}$ in this case called a "degenerate" code. Analogous types of codes can not be found in classical error correction [108].

B.3.2 Noiseless Subsystems and Decoherence Free Subspaces

Before describing the noiseless subsystem method, let us lay down some terminology. Let $\mathcal{E}$ be a quantum operation with Kraus operators $\{E_a\}$ and suppose the Hilbert space $\mathcal{H}$

factorizes as $\mathcal{H} = (\mathcal{H}_A \otimes \mathcal{H}_B) \oplus K$, with $\dim(\mathcal{H}_A) = m$, $\dim(\mathcal{H}_B) = n$, and K a subspace of arbitrary but finite dimension. Let P_{AB} be the projector onto the subspace $\mathcal{H}_A \otimes \mathcal{H}_B$, P_{kl} be projectors of the form $|\alpha_k\rangle\langle\alpha_l| \otimes \mathbb{1}_B$ for some orthonormal basis $\{|\alpha_k\rangle\} \in \mathcal{H}_A$ and the quantum operation $\mathcal{P}_{\mathcal{AB}}$ be defined by Kraus operators $\{P_{kl}\}$. The P_{kk} are called minimal reducing projections for B and $P_{AB} = \sum_{k=1}^{m} P_{kk}$ is called the minimal central projection onto $\mathcal{H}_{AB}$. Finally, let $\mathcal{S}$ be the semigroup of operators of the form $\sigma_A \otimes \sigma_B$ which are reduced by P_{AB} and have support on $P_{AB}\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.

Definition 9. *Noiseless Subsystem*

B is said to be a noiseless subsystem for $\mathcal{E}$ if $\forall\sigma_A \, \forall\sigma_B \, \exists\tau_A$

$$\mathcal{E}(\sigma_A \otimes \sigma_B) = \tau_A \otimes \sigma_B \tag{B.10}$$

Thus, B is a noiseless subsystem for $\mathcal{E}$ if there exists a quantum operation $\mathcal{F}_{\mathcal{AA}} : B(\mathcal{H}_A) \to B(\mathcal{H}_A)$ such that $\mathcal{E}|_{B(\mathcal{H}_A)\otimes B(\mathcal{H}_B)} = \mathcal{F}_{\mathcal{AA}} \otimes \mathbb{1}$. The following proposition is proved in [86]

Proposition 8. *The following four conditions are equivalent to B being a noiseless subsystem for the quantum process $\mathcal{E}$:*

1. $\forall\sigma_B \, \exists\tau_A \; : \mathcal{E}(\mathbb{1}_A \otimes \sigma_B) = \tau_A \otimes \sigma_B$
2. $\forall\sigma \in \mathcal{S} : Tr_A \circ \mathcal{P}_{\mathcal{AB}} \circ \mathcal{E}(\sigma) = Tr_A(\sigma)$
3. $\forall a : E_a$ *is invariant on* $P_{AB}\mathcal{H}$ *and* $E_a|_{P_{AB}\mathcal{H}} \in B(\mathcal{H}_A) \otimes \mathbb{1}_B$
4. $\forall a, k, l : E_a P_{AB} = P_{AB} E_a P_{AB}$ *and* $P_{kk} E_a P_{ll} = \lambda_{akl} P_{kl}$ *where* λ_{akl} *is some set of scalars.*

The noiseless subsystem framework given above encompasses the notion of decoherence free subspaces in the case when $m = 1$. This is easy to see from the first condition using $\mathcal{H}_A \otimes \mathcal{H}_B \cong \mathcal{H}_B$ and trace preservation. Hence when $m = 1$, the B subsystem is actually a subspace which is undeformed by the action of $\mathcal{E}$.

B.3.3 Unified Method For Quantum Error Correction

The unified scheme that encompasses all of these models is described by a triple $(\mathcal{R}, \mathcal{E}, \mathcal{S})$ where, as in the terminology for the standard model, $\mathcal{R}$ is a recovery quantum operation for the channel $\mathcal{E}$. $\mathcal{S}$ is a semigroup of operators defined as above in the noiseless subsystems section.

Definition 10. *Correctable Code*

For a triple $(\mathcal{R}, \mathcal{E}, \mathcal{S})$, the B subsystem is called correctable for $\mathcal{E}$ by the recovery operation $\mathcal{R}$ if it is noiseless for the quantum operation $\mathcal{R} \circ \mathcal{E}$. Concretely, using the proposition given above, B is correctable for $\mathcal{E}$ by $\mathcal{R}$ if

$$\forall \sigma_B \, \exists \tau_A \, : \mathcal{R} \circ \mathcal{E}(\mathbb{1}_A \otimes \sigma_B) = \tau_A \otimes \sigma_B \tag{B.11}$$

The standard model is encompassed within this framework in the case when $\dim(\mathcal{H}_A) = m = 1$. When $\mathcal{R} = \mathcal{I}$, B is a noiseless subsystem. If both $\mathcal{R} = \mathcal{I}$ and $m = 1$ then B is a decoherence free subspace.

The case when $\mathcal{R}$ can be chosen to be a unitary operation $\mathcal{U}$ is of particular interest and in such a scenario the subsystem B is said to be a unitarily correctable subsystem (UCS). A UCS is called a unitarily noiseless subsystem (UNS) for Λ if it is a UCS of Λ^n for all $n \geq 1$. As usual, Λ^n is the channel Λ composed with itself n times.

In the case of $\mathcal{E}$ being a unital quantum channel the following result [88] is useful for finding larger classes of codes than just noiseless subsystems,

Theorem 5. *The following are equivalent:*

1. *B is a unitarily correctable subsystem for $\mathcal{E}$*
2. *B is a noiseless subsystem for $\mathcal{E}^\dagger \circ \mathcal{E}$.*

In general, it is necessary to be able to find correctable codes for the theory to be of any practical interest. For unital channels there is an algorithm that finds all noiseless subsystems for the channel [67]. The main drawback of the algorithm is that it is exponential in the number of qubits. The key is that for unital channels the commutant and fixed point set of the channel coincide. By the Artin-Wedderburn theorem and the fact that the commutant is a finite-dimensional C^*-algebra, the matrix algebras in the Artin-Wedderburn decomposition for the commutant are areas in the Hilbert space in which noiseless quantum information may be stored. Thus the algorithm consists of how to find this decomposition of the commutant. It was proved in [30] that every noiseless subsystem for a unital channel must reside in the commutant. Hence this algorithm finds all noiseless subsystems for a unital channel. For non-unital channels, there exist algorithms for finding noiseless subsystems [30, 81].

B.4 Fault-Tolerant Quantum Computation

The theory of quantum error-correction allows for the development of a theory of fault-tolerant quantum computation which we briefly outline in this section. The general idea behind fault-tolerance comes from the following scenario: Suppose we want to implement a quantum circuit however noise affects the physical elements of the circuit (state-preparation, wires, gates and measurements). Is it possible to find a quantum-error correcting code such that concatenation of the circuit using this code reduces the error to something manageable? There are two potential significant problems with this scheme, the first being that for many error-correcting codes, the failure of a physical component in the concatenated code will lead to multiple failures on other qubits that need not even be in the same code block. Hence one needs to be wise in choosing an error-correcting code that doesn't "propagate" errors in an uncontrollable manner to other areas of the encoded circuit. Fortunately, the theory of stabilizer codes [57] provides codes where this propagation of errors can be controlled. The second potential problem is that error-correction must be performed periodically on the encoded data in order to ensure that the information is preserved. The error-correction procedures can also introduce errors into the circuit, however again if the code is chosen wisely then these errors will not propagate in an uncontrolled manner.

The brief description of fault-tolerant computing presented here will follow closely with that presented in [108]. Note that since the set of all single qubit H, S, $\frac{\pi}{8}$, coupled with $CNOT$ on all pairs of qubits, is universal we need only ensure these gates are performed "fault-tolerantly" (as well as state-preparation and measurements). Let us denote this universal set of gates on n qubits by Θ_n. We make precise the idea of a fault-tolerant operation soon. First we describe the assumed error model,

Error Model: Single qubit errors are described by Pauli errors (ie. channels whose Kraus operators are non-negative multiples of elements from $\mathcal{P}_1$) and correlated errors can occur on two qubits with these errors being described by two-qubit Pauli errors (ie. the Kraus operators are non-negative multiples of elements from $\mathcal{P}_2$).

More detailed error models have been analyzed however the basic ideas are reflected most clearly using this simple error model. We now define what it means for encoded gates, measurements and state preparations to be fault-tolerant.

Definition 11. *Fault-tolerant Encodings of Gates, Measurements and State Preparations*

A procedure for implementing an encoded gate is called fault-tolerant if the failure of any single physical component making up the encoded gate leads to at most one error

in each output encoded data block of qubits. A procedure for implementing an encoding of the measurement of an observable is said to be fault-tolerant if the failure of any single physical component making up the encoded measurement produces at most one error in each output encoded data block of qubits. In addition, letting p be the maximum probability of failure of any single component making up the measurement, if a single physical component fails the output measurement value must be correct to $1 - O(p^2)$. Lastly, a procedure for encoded state-preparation is called fault-tolerant if the failure of any single physical component produces at most one error in each output encoded data block of qubits.

The above definitions essentially state the fact that fault-tolerant operations don't allow errors to propagate in an uncontrolled manner throughout the encoded blocks of qubits. We emphasize that the property of an operation to be fault-tolerant relies directly on the particular coding scheme used and we now discuss this relationship in a bit more detail.

From these definitions one can see that a highly desirable property of the quantum code used in the procedure is that the encoded operations are applied transversally (ie. bit-wise) to the physical qubits. For instance if the Hadamard gate is encoded into an n-qubit code as $H^{\otimes n}$ then the failure of any single physical component making up the encoded Hadamard will introduce only one error on this encoded block of n qubits. The obvious question is what is the best code to use, taking into account both physical resources as well as transversality? The smallest quantum code that can correct for an arbitrary single qubit error is the 5-qubit stabilizer code [14, 89]. In terms of physical resources, this code would be best suited for fault-tolerant implementations. Unfortunately, this code does not apply many of the required encoded operations transversally. The seven qubit CSS stabilizer code [27, 135], also known as the Steane code, on the other hand does apply most of the operations transversally. Hence this code is ideally suited for fault-tolerant constructions. The encoded Clifford gates in Θ_n can be constructed transversally [108], but the $\frac{\pi}{8}$ gate can not. However as previously mentioned, $\frac{\pi}{8}$ can be simulated via a teleportation scheme by preparing an ancilla in a magic state, performing a measurement in the computational basis and using elements only from Clif_n [21]. Hence the ability to perform fault-tolerant Clifford operations, state-preparations and measurements imply the ability to fault-tolerantly implement the $\frac{\pi}{8}$ gate, and thus universal fault-tolerant quantum computation. Fault-tolerant constructions of the Clifford gates in Θ_n, as well as state preparations and measurements, using the seven qubit code are described in detail in [108].

Now that we have a coding scheme that allows for fault-tolerant quantum computation, let us focus on how fault-tolerant procedures for performing operations can actually reduce the error rates of the operations. In [108] the above definition of a fault-tolerant procedure for implementing an encoded gate is used to show that the maximum probability of intro-

ducing two or more errors into an output encoded block of qubits from the fault tolerant operation is given by cp^2 (it is explicitly proven for the encoded $CNOT$ gate). Here p is the maximum probability of failure any physical component making up the encoded operation and c is a constant that depends upon the particular coding scheme used as well as which gate we are analyzing. Assuming that the code used can correct single qubit errors, implementing a perfect recovery operation would imply the encoded procedure succeeds with probability greater than $1-cp^2$. Thus if p is small enough, the procedure for implementing the encoded operation has a smaller error rate than the physical operation itself. For the seven qubit code the constant c is roughly on the order of 10^4 for each of the operations we are interested in. Hence if $p < 10^{-4}$ there is an improvement from using the fault-tolerant operations as opposed to the physical operations themselves.

The next natural question to ask is what happens if we repeat this fault-tolerant procedure again on the "first level" encoding just described. Since the probability of failure of an operation in the first level encoded circuit is cp^2, concatenation of the fault-tolerant procedure implies that the probability of failure on the second level of encoding is $c\left(cp^2\right)^2 = c^3p^4$. Concatenating this procedure k times implies the failure probability of an element in the k'th level encoding is $\frac{(cp)^{2^k}}{c}$. As well, if the size of the physical circuit is a polynomial, $p(n)$, in the size n of some computational problem, and if R represents the maximum number of physical operations needed to implement an encoded operation in the first level encoding, then the size of the k'th level encoded circuit is bounded above by $R^k p(n)$. This concatenation of fault-tolerant procedures is the essence of the fault-tolerant threshold theorem.

Suppose one wants the overall computation to succeed with probability greater than $1-\epsilon$ for some $\epsilon > 0$. Since there are $p(n)$ $k-level$ operations in the k'th level encoding (each encoded operation contains no more than R^k physical operations), and each operation fails independently with probability at most $\frac{(cp)^{2^k}}{c}$, one will need this probability multiplied by the number of encoded gates $p(n)$ to be no larger than ϵ, ie. the requirement is

$$p(n)\frac{(cp)^{2^k}}{c} \leq \epsilon. \tag{B.12}$$

As described above, provided p is small enough (ie. $p < \frac{1}{c}$), there will exist a k for which this condition is satisfied. The requirement that p is smaller than some threshold value, denoted p_{th}, is called the "threshold condition". In the case of the seven qubit code described above, $p_{th} \sim 10^{-4}$.

Since there are $R^k p(n)$ physical gates in the k'th level encoding for which the condition in Eq. (B.12) is met. Solving for k in this condition implies that R^k is given by

$$R^k = O\left(\text{poly}\left(\frac{\log(p(n))}{\epsilon}\right)\right). \tag{B.13}$$

Hence,

$$R^k p(n) = O\left(\text{poly}\left(\frac{\log(p(n))}{\epsilon}\right) p(n)\right) \tag{B.14}$$

which is polylogarithmic in the size of the original physical circuit. Combining all of these results we can now state the fault-tolerant threshold theorem for quantum computation.

Theorem 6. *Fault-tolerant threshold theorem for quantum computation*

Let $\epsilon > 0$ and Q be a quantum circuit of size $p(n)$. Suppose the physical components of Q fail with maximum probability p under an independent, stochastic, Pauli error model. Then there exists a threshold value p_{th} such that if $p < p_{th}$, Q can be executed with probability at least $1 - \epsilon$ using

$$O\left(poly\left(\frac{\log(p(n))}{\epsilon}\right) p(n)\right) \tag{B.15}$$

physical gates.

Some important questions that arise from the above version of the threshold theorem are under what (more general) error models and fault-tolerant encoding schemes can one obtain a threshold theorem, and what are estimates of the threshold value? These are active area of research in quantum information and we give references of various results obtained in past years. Different schemes of fault-tolerant quantum computing include post-selection based methods [80], schemes based on the Bacon-Shor code [6], surface codes [121] and various CSS codes [136]. Examples of different noise models along with estimates are given by locality constraints [138], local non-Markovian noise [140], Gaussian noise [106] and long-range correlated noise [4]. Another important area of research

concerns the various architectures for fault-tolerant quantum computing, which include NMR [28], ion traps [109], semi-conductors [139], optical lattices [9], quantum dots [93] and superconducting qubits [5].

B.5 Capacities of Quantum Channels

In this section we define some capacities for transmitting various types of information through a quantum channel $\mathcal{E} : L(\mathcal{H}_A) \to L(\mathcal{H}_B)$ and give some well-known results for these capacities. For an extensive introduction to this area, and a more detailed discussion of the topics discussed here, see Ref. [70]. Intuitively speaking, the capacity of a quantum channel with respect to some type of information is the maximum achievable communication rate, where one allows for an unlimited number of uses of the channel. There are four main types of capacities of a quantum channel, distinguished by both the type of information that is being sent as well as by other resources available. These capacities are the classical ($CL(\mathcal{E})$), private classical ($PCL(\mathcal{E})$), entanglement-assisted classical ($ECL(\mathcal{E})$), and quantum ($Q(\mathcal{E})$) channel capacities. We will mainly be concerned with $CL(\mathcal{E})$ and $Q(\mathcal{E})$ and so only define and discuss these two capacities. Before discussing these quantities in more detail we define some useful entropic notions.

Definition 12. *(Von Neumann) Entropy*

The entropy of a quantum system A with Hilbert space $\mathcal{H}$ in the state $\rho \in L(\mathcal{H})$, denoted $H(A)$, is given by,

$$H(A) = -\mathrm{tr}\left(\rho \log(\rho)\right). \tag{B.16}$$

The entropy of a quantum system can be thought of as a measure of the uncertainty present in the system. Note that pure states have zero entropy and hence can be thought of as states of maximal certainty.

Definition 13. *Relative Entropy*

For density operators ρ and σ of the quantum system A, the relative entropy of ρ with respect to σ, denoted $H(\rho||\sigma)$, is given by

$$H(\rho||\sigma) = \mathrm{tr}\left(\rho \log(\rho)\right) - \mathrm{tr}\left(\rho \log(\sigma)\right) \tag{B.17}$$

where if $\rho|_{ker(\sigma)} \neq 0$,

$$H\left(\rho||\sigma\right) := \infty. \tag{B.18}$$

It can be shown that the relative entropy is non-negative and equal to 0 if and only if $\rho = \sigma$.

Definition 14. *Joint and Conditional Entropy*

For a composite quantum system AB (with Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$*) in the state* ρ_{AB}*, the joint entropy,* $H\left(A, B\right)$*, is*

$$H(A, B) = -\mathrm{tr}\left(\rho_{AB} \log(\rho_{AB})\right). \tag{B.19}$$

The conditional entropy of A with respect to B, denoted $H(A|B)$*, is given by*

$$H(A|B) = H(A, B) - H(B). \tag{B.20}$$

An interesting property of the conditional entropy for quantum systems is that it can be negative (which is not possible for the classical Shannon conditional entropy).

Definition 15. *Mutual Information*

The mutual information of a composite quantum system AB in the state ρ_{AB} *is denoted* $I(A : B)$ *and is given by*

$$I(A : B) = H\left(A\right) + H\left(B\right) - H\left(A, B\right). \tag{B.21}$$

The mutual information can be thought of as the amount of information that is common to both systems. Loosely speaking, an intuitive method for remembering the above formula comes from drawing an analogy to Venn diagrams for sets.

We now define one of the most important concepts in the theory of capacities of quantum channels, the Holevo χ-quantity [68]. First we define the χ-quantity for an ensemble of states.

Definition 16. *Holevo χ-Quantity (Holevo Information) of an Ensemble*

For an ensemble $\{p_x, \rho_x\}$, $x \in X$, the Holevo χ-quantity (aka Holevo information) of $\mathcal{E}$ with respect to this ensemble is defined to be

$$\chi\left(\mathcal{E}\right)_{\{p_x,\rho_x\}} = H\left(\sum_x p_x \mathcal{E}\left(\rho_x\right)\right) - \sum_x p_x H\left(\mathcal{E}\left(\rho_x\right)\right). \tag{B.22}$$

An alternative expression for the Holevo χ-quantity of $\{p_x, \rho_x\}$, $x \in X$, can be obtained by noting that if the state of the composite space XB is given by

$$\sigma_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right) \in L\left(\mathbb{C}^{|X|} \otimes \mathcal{H}_B\right) \tag{B.23}$$

then,

$$\chi(\mathcal{E})_{\{p_x,\rho_x\}} = I(X:B)_{\sigma_{XB}}. \tag{B.24}$$

To see this note that

$$H\left(\sigma_X\right) = H(\{p_x\}), \tag{B.25}$$

$$H\left(\sigma_B\right) = H\left(\sum_x p_x \mathcal{E}\left(\rho_x\right)\right), \tag{B.26}$$

and

$$
\begin{aligned}
H\left(\sigma_{XB}\right) &= H\left(\sum_x p_x|x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right)\right) \\
&= -\mathrm{tr}\left(\left(\sum_x p_x|x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right)\right)\log\left(\sum_x p_x|x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right)\right)\right) \\
&= -\mathrm{tr}\left(\left(\sum_x p_x|x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right)\right)\left(\sum_x |x\rangle\langle x| \otimes \log\left(p_x\mathcal{E}\left(\rho_x\right)\right)\right)\right) \\
&= -\mathrm{tr}\left(\left(\sum_x p_x|x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right)\right)\left(\sum_x |x\rangle\langle x| \otimes \left(\log(p_x)\mathbb{1}_B + \log\left(\mathcal{E}\left(\rho_x\right)\right)\right)\right)\right) \\
&= -\mathrm{tr}\left(\sum_x p_x \log(p_x)|x\rangle\langle x| \otimes \mathcal{E}\left(\rho_x\right)\right) - \mathrm{tr}\left(\left(\sum_x p_x|x\rangle\langle x|\right) \otimes \mathcal{E}\left(\rho_x\right)\log\left(\mathcal{E}\left(\rho_x\right)\right)\right) \\
&= H(\{p_x\}) + \sum_x p_x H(\mathcal{E}\left(\rho_x\right)). \qquad \text{(B.27)}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
I(X:B) &= H\left(\sigma_X\right) + H\left(\sigma_B\right) - H\left(\sigma_{XB}\right) \\
&= H(\{p_x\}) + H\left(\sum_x p_x\rho_x\right) - H(\{p_x\}) - \sum_x p_x H(\rho_x) \\
&= H\left(\sum_x p_x\rho_x\right) - \sum_x p_x H(\rho_x). \qquad \text{(B.28)}
\end{aligned}
$$

Definition 17. *Holevo χ-Quantity (Information) of a Quantum Channel*
The Holevo χ-quantity of $\mathcal{E}$, denoted $\chi(\mathcal{E})$, is given by,

$$
\begin{aligned}
\chi(\mathcal{E}) &= max_{\{p_x,\rho_x\}}\chi(\mathcal{E})_{\{p_x,\rho_x\}} \\
&= max_{\{p_x,\rho_x\}}I(X:B)_{\sigma_{XB}}. \qquad \text{(B.29)}
\end{aligned}
$$

We now define the coherent information of a quantum operation by first defining the

coherent information with respect to a particular state.

Definition 18. *Coherent Information of $\mathcal{E}$ With Respect to ρ*

Let A be a quantum system in the state ρ and $\mathcal{E}$ be a quantum operation on A such that $(E,\mathcal{U})$ is a Stinespring representation for $\mathcal{E}$, ie. $\mathcal{E}(\rho) = \mathrm{tr}_E\left(U\rho\otimes|0\rangle\langle 0|U^\dagger\right)$. Let $\sigma = U\rho\otimes|0\rangle\langle 0|U^\dagger \in L\left(\mathcal{H}_A\otimes\mathcal{H}_E\right)$ and define the complimentary channel of $\mathcal{E}$, $\mathcal{E}_C$, by

$$\mathcal{E}_C(\rho) = \mathrm{tr}_A(\sigma). \tag{B.30}$$

The coherent information of $\mathcal{E}$ with respect to ρ, denoted $I_{coh}(\mathcal{E})_\rho$, is given by

$$\begin{aligned} I_{coh}(\mathcal{E})_\rho &= H(A)_\sigma - H(E)_\sigma \\ &= H\left(\mathcal{E}(\rho)\right) - H\left(\mathcal{E}_C(\rho)\right). \end{aligned} \tag{B.31}$$

Definition 19. *Coherent Information of $\mathcal{E}$*

The coherent information of $\mathcal{E}$ is given by the maximum of $I_{coh}(\mathcal{E})_\rho$ over all input states ρ,

$$I_{coh}(\mathcal{E}) = max_\rho\left(H(A)_\sigma - H(E)_\sigma\right). \tag{B.32}$$

$H\left(\mathcal{E}_C(\rho)\right)$ is commonly called the "exchange entropy" and is denoted $H_e\left(\rho,\mathcal{E}\right)$. Hence writing $H(\rho,\mathcal{E}) = H(\mathcal{E}(\rho))$ we have,

$$I_{\mathrm{coh}}(\mathcal{E}) = \mathrm{max}_\rho\left(H(\rho,\mathcal{E}) - H_e\left(\rho,\mathcal{E}\right)\right). \tag{B.33}$$

In the case that $I_{\mathrm{coh}}(\mathcal{E})$ is negative we set $I_{\mathrm{coh}}(\mathcal{E}) = 0$.

We are now ready to define the classical and quantum capacities of a quantum channel; first we deal with the classical capacity. The definitions of capacity rely on the idea of "achievable rates" which we define separately for a more lucid presentation. Intuitively, a "rate" for a channel corresponds to a particular number of messages that can be sent per some number of uses of the channel.

Definition 20. *ϵ-Classical-Achievable Rate of a Quantum Channel*

Suppose Alice and Bob have agreed on an alphabet $\Sigma = \{1, ..., |\Sigma|\}$ for their classical communication and Alice wants to send a message to Bob using both the alphabet and a noisy quantum channel $\mathcal{E} : L(\mathcal{H}_A) \to L(\mathcal{H}_B)$. A message m will consist of a particular string of elements from Σ. Let the set of all possible messages be denoted $\mathcal{M}$. $\mathcal{M}$ can be infinite but need not contain all possible finite strings of elements from Σ (as in the case of the english alphabet and the language formed from it).

Let $\epsilon > 0$. A real number R is called an ϵ-classical-achievable rate for $\mathcal{E}$ if there exists $n_\epsilon \in \mathbb{N}$ such that for every $n \geq n_\epsilon$ the following is satisfied;

1. Alice chooses a classical code $\{\pi_k, \rho_k \in L\left(\mathcal{H}_A^{\otimes n}\right)\}_{k=1}^{K_n}$ (ie. finite set of quantum states with some probability distribution $\{\pi_k\}_{k=1}^{K_n}$) such that $\frac{\log(K_n)}{n} \geq R$ and Bob chooses a decoding operation $\mathcal{D}_n$ represented by a POVM measurement with POVM elements E_m^n, $m \in \{1, ..., K_n\}$.

The code can be thought of as representing a probability distribution π on K_n messages from $\mathcal{M}$ represented by $\eta = \{\eta_1, ..., \eta_{K_n}\}$ (although such a representation is not necessary; the important point is the probability distribution π on the K_n states with $\frac{\log(K_n)}{n} \geq R$). Thus there is a bijection between η and $\{\rho_k \in L\left(\mathcal{H}_A^{\otimes n}\right)\}_{k=1}^{K_n}$ given simply by $\eta_i \to \rho_i$.

2. For any message η_i Alice chooses from η according to π, she represents η_i via the corresponding element ρ_i in the code space, and sends it through $\mathcal{E}^{\otimes n}$ to Bob.

3. Bob performs the POVM measurement of $\mathcal{E}^{\otimes n}(\rho_i)$. Let M be the random variable for the message η_i Alice chooses and M' represents Bob's random variable of the output of the measurement. Then,

$$pr\{M' = \eta_i | M = \eta_i\} = \mathrm{tr}\left(E_i^n \mathcal{E}^{\otimes n}(\rho_i)\right). \tag{B.34}$$

Thus the probability of error by Bob's measurement is $1 - pr\{M' = i | M = i\} = \mathrm{tr}\left((\mathbb{1} - E_i^n)\mathcal{E}^{\otimes n}(\rho_i)\right)$ and the probability of error by Bob's measurement taking into account the coding distribution π, denoted $p_e(m)$, is,

$$\begin{aligned} p_e(i) &= (1 - pr\{M' = i | M = i\})\,\pi_i & \text{(B.35)} \\ &= \mathrm{tr}\left((\mathbb{1} - E_i^n)\mathcal{E}^{\otimes n}(\rho_i)\right)\pi_i. & \text{(B.36)} \end{aligned}$$

4. $\|\vec{p_e}\|_\infty = max_m p_e(m) \leq \epsilon$.

It is important to note in the above definition that for each $n \geq n_\epsilon$, the only freedom Alice and Bob have comes from the coding scheme $\{\pi_k, \rho_k \in L\left(\mathcal{H}_A^{\otimes n}\right)\}_{k=1}^{K_n}$ and the decoding POVM with elements E_m^n, $m \in \{1, ..., K_n\}$. Hence one can say that R is ϵ-classical-achievable if there exists n_ϵ such that for every $n \geq n_\epsilon$ there exists a classical code and decoding POVM, both of cardinality K_n, such that both $\log(K_n) \geq nR$ and $\|\vec{p_e}\|_\infty \leq \epsilon$ are satisfied for $\vec{p_e}$ defined above. In more plain terms, there exists n_ϵ such that for every $n \geq n_\epsilon$, Alice can form a number of messages K_n such that $K_n \geq e^{Rn}$ and no matter what message Alice chooses, Bob can decode it with error at most ϵ.

Definition 21. *Classical-Achievable Rate of a Quantum Channel*

R is said to be a classical-achievable rate of the quantum channel $\mathcal{E}$ if for every $\epsilon > 0$, R is an ϵ-classical-achievable rate.

Definition 22. *Classical Capacity of a Quantum Channel*

The classical capacity of the quantum channel $\mathcal{E}$, $CL(\mathcal{E})$, is defined to be the supremum over all classical-achievable rates R.

We note that if Alice is restricted to using only product states (although they can be mixed and entangled entangled within each subsystem) then the capacity in this restricted setting, denoted $CL^1(\mathcal{E})$, is called the product state classical capacity. The HSW theorem [69, 127] gives a useful characterization of $CL^1(\mathcal{E})$ in terms of the Holevo χ-quantity of $\mathcal{E}$, $\chi(\mathcal{E})$.

Theorem 7. *HSW Theorem*

For a quantum channel $\mathcal{E}$,

$$CL^1(\mathcal{E}) = \chi(\mathcal{E}) \tag{B.37}$$

and,

$$\begin{aligned} CL(\mathcal{E}) &= \lim_{n\to\infty} \frac{1}{n} CL^1(\mathcal{E}^{\otimes n}) \\ &= \lim_{n\to\infty} \frac{1}{n} \chi\left(\mathcal{E}^{\otimes n}\right). \end{aligned} \tag{B.38}$$

An obvious corollary of Eq. (B.37) is that,

$$CL(\mathcal{E}) \geq \chi(\mathcal{E}) \tag{B.39}$$

since restricting to product states will put a lower bound on the amount of information that can be sent through a quantum channel.

For a function f on quantum channels, the limit $\lim_{n\to\infty} \frac{1}{n} f\left(\mathcal{E}^{\otimes n}\right)$ is called the "regularization" of f. Hence Eq. (B.38) shows that the classical capacity of $\mathcal{E}$ is equal to the regularization of the product state classical capacity. A function f on quantum channels is called additive if $f\left(\mathcal{E}_1 \otimes \mathcal{E}_2\right) = f\left(\mathcal{E}_1\right) + f\left(\mathcal{E}_2\right)$. One can see from Eq. (B.37) that if the Holevo χ-quantity is additive then in fact the classical capacity is equal to the product state classical capacity (and also equal to χ). This question is equivalent to various other additivity questions in quantum information theory, for example the minimum output entropy and entanglement of formation [133]. It was widely believed for some time that χ is additive however it has recently been shown that this is false [64] and so in general $CL(\mathcal{E}) \neq CL^1(\mathcal{E})$. With the tools we have developed in discussing the classical capacity we can now define the quantum capacity of a quantum channel.

The sequence of definitions leading to the quantum capacity of a quantum channel is similar, and in many ways simpler, than that of the classical capacity.

Definition 23. *ϵ-Achievable Rate of a Quantum Channel*

Suppose Alice wants to send quantum states to Bob using the noisy channel $\mathcal{E} : L\left(\mathcal{H}_A\right) \to L\left(\mathcal{H}_B\right)$. Let $\epsilon > 0$. A real number R is called an ϵ-classical-achievable rate for $\mathcal{E}$ if there exists $n_\epsilon \in \mathbb{N}$ such that for every $n \geq n_\epsilon$ the following is satisfied;

1. Alice chooses a subspace C_n of $\mathcal{H}_A^{\otimes n}$ (called the code subspace) with dimension K_n such that $\frac{\log(K_n)}{n} \geq R$ and Bob chooses a decoding operation $\mathcal{D}_n : L\left(\mathcal{H}_B^{\otimes n}\right) \to L\left(C_n\right)$. The decoding operation is allowed to be a general quantum operation.

2. For an arbitrary pure state $\rho = |\psi\rangle\langle\psi|$ with support in C_n (ie. $\rho \in L\left(C_n\right)$), Alice sends ρ through $\mathcal{E}^{\otimes n}$ to Bob.

3. Bob applies $\mathcal{D}_n$ to $\mathcal{E}^{\otimes n}\left(\rho\right)$ to obtain $\mathcal{D}_n \circ \mathcal{E}^{\otimes n}\left(\rho\right)$.

4. The fidelity between $\mathcal{D}_n \circ \mathcal{E}^{\otimes n}\left(\rho\right)$ and ρ satisfies,

$$F\left(\mathcal{D}_n \circ \mathcal{E}^{\otimes n}\left(\rho\right), \rho\right) \geq 1 - \epsilon. \tag{B.40}$$

As for the classical capacity, it is important to note in the above definition that for each $n \geq n_\epsilon$, the only freedom Alice and Bob have comes from the choice of code space C_n and decoding operation $\mathcal{D}_n$. Hence one can say that R is ϵ-achievable if there exists n_ϵ such that for every $n \geq n_\epsilon$ there exists a code subspace $C_n \subseteq \mathcal{H}_A^{\otimes n}$ and decoding operation $\mathcal{D}_n : L\left(\mathcal{H}_B^{\otimes n}\right) \rightarrow L\left(C_n\right)$ such that $\log\left(K_n\right) \geq nR$ and for any $\rho = |\psi\rangle\langle\psi| \in C_n$, $F\left(\mathcal{D}_n \circ \mathcal{E}^{\otimes n}\left(\rho\right), \rho\right) \geq 1 - \epsilon$. Note that the reason we can restrict attention to pure states is by the concavity of the fidelity.

Definition 24. *Achievable Rate of a Quantum Channel*

R is said to be an achievable rate of the quantum channel $\mathcal{E}$ if for every $\epsilon > 0$, R is an ϵ-achievable rate.

Definition 25. *Quantum Capacity of a Quantum Channel*

The quantum capacity of the quantum channel $\mathcal{E}$, $Q(\mathcal{E})$, is defined to be the supremum over all achievable rates R.

As in the case of the classical capacity, there are useful characterizations of $Q(\mathcal{E})$ [95, 130, 44],

Theorem 8. *For a quantum channel $\mathcal{E}$,*

$$Q(\mathcal{E}) \geq I^{coh}(\mathcal{E}) \tag{B.41}$$

and regularizing the coherent information gives,

$$Q(\mathcal{E}) = \lim_{n\rightarrow\infty} \frac{1}{n} I^{coh}(\mathcal{E}^{\otimes n}). \tag{B.42}$$

The coherent information is known to be non-additive and so in general, $Q(\mathcal{E}) \neq I^{\mathrm{coh}}(\mathcal{E})$.

Appendix C

Unitary t-Designs and Twirling Quantum Channels

In this chapter we introduce unitary t-designs and prove some basic results about them. Unitary t-designs naturally lead to the concept of twirling a quantum channel over a subset of the unitary group $U(d)$. We will discuss twirling over all of $U(d)$ using the Haar measure and then look at applications that involve estimating the average gate fidelity of a quantum gate. In the next chapter we will discuss twirling over discrete subsets of the unitary group.

C.1 Unitary t-Designs

Before defining unitary t-designs we discuss the well-known concept in numerical analysis of spherical t-designs [42]. Suppose one has a function defined on the unit sphere $S^{n-1} \subseteq \mathbb{R}^n$ and wants to compute the average of the function. The maximal symmetry of the domain suggests that for "well-behaved" classes of functions there should exist a set of fixed points on $\mathbb{S}^{n-1}$ such that for any function f in the class, the average of the values of f at these points is equal to the global average of f on $\mathbb{S}^{n-1}$. Polynomials are an important class of functions in mathematics and specifically numerical analysis, and are divided into a countable number of classes by their degree. A natural class of functions to analyze the existence of such a set of points is the set of polynomials of degree t. These sets are called spherical t-designs. Formally, this is stated below

Definition 26. *Spherical t-Design*

A spherical t-design is a finite set of points $\{x_1, ..., x_K\} \subseteq \mathbb{S}^{n-1}$ *such that for any polynomial* $p : \mathbb{S}^{n-1} \to \mathbb{R}$ *of degree less than or equal to t, the average of p over* $\mathbb{S}^{n-1}$ *with respect to the rotationally-invariant Haar measure is equal to the average of the polynomial values at each* x_i*. The polynomials defined on* $\mathbb{S}^{n-1}$ *are just the set of all polynomials of degree less than or equal to t defined on* $\mathbb{R}^n$*, but restricted to* $\mathbb{S}^{n-1}$*.*

In the case of $\mathbb{S}^2$ we require that for any polynomial p of degree less than or equal to t whose domain is $\mathbb{R}^3$,

$$\int_0^{2\pi} \int_0^{\pi} p(\theta, \phi) sin\theta d\theta d\phi = \frac{1}{K} \sum_{j=1}^{K} p(x_j) \tag{C.1}$$

where we utilize the usual spherical coordinate system on the sphere. A specific example is given by the 3-design for $\mathbb{S}^2$ where $K = 6$ and the x_j are chosen so that they form the vertices of a regular octahedron. It has been proved [128] that spherical t-designs exist of sufficiently large sizes. More precisely, there exists a number $N(n, t)$ such that $\forall$N $\geq$ N(n,t) there exists a spherical t-design of N points on $\mathbb{S}^n$. Only estimates of the size of $N(n, t)$exist.

A unitary t-design is similar in principle to that of a spherical t-design. Let $\mathcal{H} = \mathbb{C}^d$ and recall that a homogeneous polynomial is one in which all the monomials making up the polynomial have the same degree. The definition of a unitary t-design is as follows [40],

Definition 27. *Unitary t-Design*

A unitary t-design is a finite set $\{U_1, ..., U_K\} \subseteq U(d)$ *of unitary matrices such that for every homogeneous complex-valued polynomial p in* $2d^2$ *indeterminates of degree (s,s) less than or equal to (t,t),*

$$\frac{1}{K} \sum_{j=1}^{K} p(U_j) = \int_{U(d)} p(U) d\mu_H(U). \tag{C.2}$$

The integral is taken with respect to the unique bi-invariant normalized Haar measure μ_H on $U(d)$ (see Sec.D.1.2). $p(U)$ is defined to be the evaluation of p at the $2d^2$ matrix entries, and their complex conjugates, of U. That is, without loss of generality, if the indeterminates are labelled $x_1, ..., x_{2d^2}$ we can relabel the them by the mapping.

$$
\begin{aligned}
x_1 &\rightarrow U_{1,1} \\
x_2 &\rightarrow U_{1,2} \\
&\cdot \\
&\cdot \\
x_{d^2} &\rightarrow U_{d,d} \\
x_{d^2+1} &\rightarrow \overline{U_{1,1}} \\
x_{d^2+2} &\rightarrow \overline{U_{1,2}} \\
&\cdot \\
&\cdot \\
x_{2d^2} &\rightarrow \overline{U_{d,d}}
\end{aligned} \tag{C.3}
$$

Then the evaluation of p comes from choosing a specific $U \in U(d)$ as in the definition. Under this association, p having degree equal to (s, s) means that each monomial has $2s$ indeterminates where s of them are from the set $\{U_{1,1}, ..., U_{d,d}\}$ and the remaining s must come from $\{\overline{U_{1,1}}, ..., \overline{U_{d,d}}\}$. The following gives an equivalent characterization of a unitary t-design. The proof is an obvious extension of Corollary 5.2.2 in [40]. We give it here for completeness.

Proposition 9. *$\{U_1, ..., U_K\}$ is a unitary t-design if and only if $\forall s \in \{0, ..., t\}$, $\forall m, n \in \{1, ..., d\}$ and $\forall \rho \in B(\mathcal{H}^{\otimes s})$,*

$$
\frac{1}{K} \sum_{j=1}^{K} P_{m,n} \left(U_j^{\otimes s} \rho U_j^{\otimes s \dagger} \right) = \int_{U(d)} P_{m,n} \left(U^{\otimes s} \rho U^{\otimes s \dagger} \right) d\mu_H(U). \tag{C.4}
$$

Here we have denoted the s-fold tensor product of an operator A with itself by $A^{\otimes s}$, and $P_{m,n}$ corresponds to the projector onto the (m, n) entry of a matrix.

Proof. First, suppose that $\{U_j\}_{j=1}^K$ form a unitary t-design. Note that the entries of $U^{\otimes s}$ are just the set of all monomials of degree s evaluated at the matrix entries of U. Similarly, the entries of $U^{\otimes s \dagger}$ are just the set of all monomials of degree s evaluated at the conjugates of the matrix entries of U. Thus, the matrix entries of $U^{\otimes s} \rho U^{\otimes s \dagger}$ are homogeneous degree (s, s) polynomials in the $2d^2$ indeterminates given by the entries of U and $U^\dagger$. This shows that for each $m, n \in \{1, ..., d\}$,

$$\frac{1}{K}\sum_{j=1}^{K} P_{m,n}\left(U_j^{\otimes s}\rho U_j^{\otimes s\dagger}\right) = \int_{U(d)} P_{m,n}\left(U^{\otimes t}\rho U^{\otimes t\dagger}\right) d\mu_H(U). \tag{C.5}$$

The converse is also simple. Note that every Hermitian matrix is a (real) linear combination of states. Hence, if

$$\frac{1}{K}\sum_{j=1}^{K} P_{m,n}\left(U_j^{\otimes s}\rho U_j^{\otimes s\dagger}\right) = \int_{U(d)} P_{m,n}\left(U^{\otimes s}\rho U^{\otimes s\dagger}\right) d\mu_H(U) \tag{C.6}$$

holds for all states ρ, then it holds for every hermitian matrix. The fact that there exists a Hermitian basis for $B(\mathcal{H}^{\otimes s})$ implies that the statement holds for any linear operator $A \in B(\mathcal{H}^{\otimes s})$. Finally, any monomial of degree (s,s) in $2d^2$ indeterminates can be constructed in one of the d^2 entries of $\left(U^{\otimes s}AU^{\otimes s\dagger}\right)$ by choosing A appropriately. By linearity, the definition for a unitary t-design is satisfied.

□

It can be seen in a manner similar to the above proof that the condition

$$\frac{1}{K}\sum_{j=1}^{K} P_{m,n}\left(U_j^\dagger M_1 U_j M_2 ... U_j^\dagger M_{2s-1} U_j\right) = \int_{U(d)} P_{m,n}\left(U_j^\dagger M_1 U_j M_2 ... U_j^\dagger M_{2s-1} U_j\right) d\mu_H(U) \tag{C.7}$$

holding for all $s \in \{0,...,t\}$, all $m,n \in \{1,...,d\}$ and all linear operators $M_1,..,M_{2s-1}$ is equivalent to a t-design. Indeed, the definition of a t-design clearly implies the above and conversely any monomial of degree (s,s) can be constructed by appropriately choosing the M_1 through M_{2s-1}.

For $d = 2^n$, exact unitary t-designs have been constructed for $t = 1$ and $t = 2$ [40]. The Clifford group forms a unitary 2-design while the Pauli group forms a 1-design. It is unknown whether there exist unitary t-designs for $t \geq 3$. For the specific case of $t = 2$ it can be shown from above that $\{U_1,...,U_K\}$ satisfying the condition for a unitary 2-design is equivalent to

$$\frac{1}{K}\sum_{j=1}^{K} P_{m,n}\left(U_j\Lambda\left(U_j^\dagger\rho U_j\right)U_j^\dagger\right) = \int_{U(d)} P_{m,n}\left(U\Lambda\left(U^\dagger\rho U\right)U^\dagger\right) d\mu_H(U) \tag{C.8}$$

being satisfied for any quantum channel Λ and any state ρ [40]. This naturally leads to the concept of twirling.

C.2 Twirling Quantum Channels

Twirling a quantum channel Λ over $U(d)$ consists of averaging Λ under the composition $\mathcal{U} \circ \Lambda \circ \mathcal{U}^\dagger$ where the unitary operations $\mathcal{U}(\rho) = U\rho U^\dagger$ are chosen according to some probability measure μ [14, 40]. The averaged channel

$$\begin{aligned} \bar{\Lambda}(\rho) &= \int_{U(d)} \mathcal{U} \circ \Lambda \circ \mathcal{U}^\dagger(\rho) d\mu(\mathcal{U}) \\ &= \int_{U(d)} U\Lambda(U^\dagger \rho U)U^\dagger d\mu(U) \end{aligned} \tag{C.9}$$

is known as the "twirled channel". The case where the distribution over unitaries is discrete is of practical interest. In this case, the twirled channel is written as,

$$\bar{\Lambda}(\rho) = \sum_i \mathrm{pr}(\mathcal{U}_i)\, \mathcal{U}_i \circ \Lambda \circ \mathcal{U}_i^\dagger(\rho) \tag{C.10}$$

where $\{\mathrm{pr}(\mathcal{U}_i)\}$ is a probability distribution over the $\mathcal{U}_i$.

Hence, from the previous section we have the following proposition,

Proposition 10. *$\{U_1, ..., U_K\}$ forms a unitary 2-design if and only if for any quantum channel Λ, the uniform twirl of Λ over $\{U_1, ..., U_K\}$ is equal to the full Haar twirl of Λ.*

Since a uniform probability distribution on the Clifford group is a unitary 2-design [14, 40], twirling a channel over the Clifford group produces the same result as twirling over $U(d)$. More precisely, if $\mathrm{Clif}_n = \{\mathcal{C}_j : j \in \mathcal{K} = \{1, ..., |\mathrm{Clif}_n|\}\}$ then,

$$\begin{aligned} \mathcal{W}(\Lambda)(\rho) &:= \frac{1}{|\mathrm{Clif}_n|} \sum_{j=1}^{|\mathrm{Clif}_n|} \left(C_j \Lambda \left(C_j^\dagger \rho C_j \right) C_j^\dagger \right) \\ &= \int_{U(d)} \left(U\Lambda \left(U^\dagger \rho U \right) U^\dagger \right) dU. \end{aligned} \tag{C.11}$$

This property is extremely useful in various areas of quantum information theory, one example being noise characterization. Indeed, this is used extensively in the randomized benchmarking protocol presented in this thesis and outlined in [99]. As shown in [107, 46], $\int_{U(d)} \left(U\Lambda\left(U^\dagger \rho U\right)U^\dagger\right) dU$ produces the unique depolarizing channel Λ_d with the same average fidelity as Λ. Hence if $\overline{\mathcal{F}_{\Lambda,\mathcal{I}}}$ is the average fidelity of Λ, and Λ_d is given by

$$\Lambda_d(\rho) = p\rho + (1-p)\frac{\mathbb{1}}{d} \tag{C.12}$$

then,

$$\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} = p + \frac{(1-p)}{d}. \tag{C.13}$$

Thus twirling a quantum operation over the Clifford group produces a depolarizing channel and the average fidelity is invariant under the twirling operation.

In randomized benchmarking we are concerned with compositions of both gate-independent and gate-dependent twirls. In the gate-independent case, the sequence of twirls of Λ of length k, $\mathcal{W}(\Lambda)^k$, can be re-written as the k-fold composition of Λ_d with itself. Using the above representation of Λ_d we get,

$$\mathcal{W}(\Lambda)^k(\rho) = p^k\rho + (1-p^k)\frac{\mathbb{1}}{d}. \tag{C.14}$$

Therefore the average fidelity decreases exponentially to $\frac{1}{d}$ since,

$$\overline{\mathcal{F}_{\Lambda_d^k,\mathcal{I}}} = p^k + \frac{(1-p^k)}{d}. \tag{C.15}$$

We can also write the average fidelity of Λ in terms of its χ-matrix written with respect to the Pauli basis (see Sec. A.2). As shown in [33],

$$\overline{\mathcal{F}_{\Lambda,\mathcal{I}}} = \frac{\chi_{0,0}d + 1}{d+1} \tag{C.16}$$

which gives,

$$\chi_{0,0} = p\left(1 - \frac{1}{d^2}\right) + \frac{1}{d^2} = \frac{\overline{\mathcal{F}_{\Lambda,\mathcal{I}}}(d+1) - 1}{d}. \tag{C.17}$$

Therefore $\chi_{0,0}$ for a quantum operation (written with respect to the Pauli basis) is invariant under twirling over a 2-design. Moreover $\chi_{0,0}$ for Λ_d^k decreases to 0 exponentially in k.

C.3 Permutation Operators and the Symmetric Subspace

In Sec. 3 we will rely heavily on the theory of permutation operators, symmetric subspaces and computing averages over Haar measures. To compute averages over a unitarily invariant measure we will begin by transforming polynomial functions of degree k into linear functions on k copies of the Hilbert space $\mathcal{H}$. We will then rely on a simple and beautiful result called *Schur-Weyl duality*, which states (in essence) that the actions of the unitary group and the permutation group (on such a k-fold tensor product) commute, and their irreducible representations (irreps) share a set of labels. Rather than discuss Schur-Weyl duality in detail, we will only introduce the tools that we need. In this section, we will briefly discuss permutation operators, the symmetric group, the totally symmetric subspace of $\mathcal{H}^{\otimes k}$, and a couple of technical results that will be of use.

Let $\mathcal{H}$ be a Hilbert space and $\mathcal{H}^{\otimes k}$ a tensor product of k copies of it. If S_k is the symmetric group on k objects and $\sigma \in S_k$ is a permutation, then there exists a unitary operator $\mathcal{P}_\sigma$ that implements σ on $\mathcal{H}^{\otimes k}$:

$$\mathcal{P}_\sigma \left(|\psi_1\rangle \otimes ... \otimes |\psi_k\rangle \right) = \left|\psi_{\sigma^{-1}(1)}\right\rangle \otimes ... \otimes \left|\psi_{\sigma^{-1}(k)}\right\rangle .$$

The totally symmetric subspace of $\mathcal{H}^{\otimes k}$ comprises all the states that are invariant under every such permutation operator – or, to put it another way, it is the intersection of the $+1$ eigenspaces of all $\mathcal{P}_\sigma$. The projector onto this space is given by

$$\pi_{\text{sym}}(k, d) = \frac{1}{k!} \sum_{\sigma \in S_k} \mathcal{P}_\sigma.$$

This projector appears in integrals over the unitary group, for the following reason (see Ref. [122]). Suppose we take a state $|\psi\rangle \in \mathcal{H}$, and then construct the projector onto its k-fold tensor product, $|\psi\rangle\langle\psi|^{\otimes k}$. This projector is a $+1$ eigenoperator of every permutation, so it lies in the totally symmetric subspace. Now, if we take the *average* of all such projectors according to the Fubini-Study Haar measure μ_{FS} on pure states (denoted $\overline{|\psi\rangle\langle\psi|^{\otimes k}}$), then we get an operator in $L\left(\mathcal{H}^{\otimes k}\right)$ that: (i) is invariant under all unitaries $U^{\otimes k}$; (ii) is supported on the totally symmetric subspace; and (iii) has unit trace. By Schur's Lemma, a unitarily

invariant operator is a weighted sum of projectors onto irreducible representations of the unitary group. The only such operators supported on the totally symmetric subspace are proportional to $\pi_{\rm sym}$ itself. Since $\overline{|\psi\rangle\langle\psi|^{\otimes k}}$ has unit trace,

$$\overline{|\psi\rangle\langle\psi|^{\otimes k}} \equiv \int_{\psi \in CP^{d-1}} |\psi\rangle\langle\psi|^{\otimes k} d\mu_{FS} = \frac{\pi_{\rm sym}(k,d)}{{\rm tr}\left[\pi_{\rm sym}(k,d)\right]}. \tag{C.18}$$

The normalization constant is easy to evaluate by counting arguments. The symmetric subspace of $\mathcal{H}^{\otimes k}$ is spanned by the bosonic Fock states, $|n_1, n_2, \ldots, n_d\rangle$, which are indexed by the number of particles n_i in state i, subject to $\sum_i n_i = k$. Counting such states, we get

$$\begin{aligned} {\rm tr}\left[\pi_{\rm sym}(k,d)\right] &= \binom{k+d-1}{d-1} \\ &= \frac{d(d+1)(d+2)\ldots(d+k-1)}{k!}. \end{aligned} \tag{C.19}$$

Suppose that we have k operators $A_1, ..., A_k$ in $L(\mathcal{H})$, and a permutation $\sigma \in S_k$ written as a product of disjoint cycles $(a_1 ... a_r) ... (a_q ... a_k)$. Then

$${\rm tr}\left[(A_1 \otimes ... \otimes A_k)\ \mathcal{P}_\sigma\right] = {\rm tr}\left[A_{a_1} ... A_{a_r}\right] ... {\rm tr}\left[A_{a_q} ... A_{a_k}\right]. \tag{C.20}$$

So, to calculate ${\rm tr}\left[(A_1 \otimes ... \otimes A_k)\ \mathcal{P}_\sigma\right]$, we can write σ in cyclic notation, replace "i" with operator A_i, and replace each "( )" with "tr[ ]".

Appendix D

Concentration of Measure

D.1 Topology and Measure Theory

D.1.1 Topology

The following is a basic introduction to topology. A reference for further reading is [105].

Definition 28. *Topology*

A topology τ on a set X is a subset of the power set, $\mathcal{P}(X)$, of X that satisfies

1. *If A_i, $i \in I$, are in τ then $\cup_{i \in I} A_i \in \tau$,*
2. *If A and B are in τ then $A \cap B$ is in τ,*
3. *$\emptyset$, $X \in \tau$.*

Elements of τ are called open sets. A set X with a topology τ defined on it is called a topological space and denoted (X, τ).

Definition 29. *Open Cover, Compact Topological Space*

If X is a topological space then an open cover of X is a subset A_i of τ such that $\cup_{i \in I} A_i = X$. A compact topological space is one for which every open cover contains a finite sub-collection that also covers X.

If X and Y are topological spaces with topologies τ_1 and τ_2, there is a natural topology, τ_p, one can put on $X \times Y$ called the product topology. Elements of τ_p are arbitrary unions of sets of the form $U \times V$ where $U \in \tau_1$ and $V \in \tau_2$.

Definition 30. *Continuous Function*

A function f from a topological space (X, τ_1) to a topological space (Y, τ_2) is called continuous if for every open set V in τ_2, $f^{-1}(V)$ is open in τ_1.

Definition 31. *Topological Group*

A topological group is both a group and a topological space (X, τ) such that

1. *The group operation is continuous from $(X \times X, \tau_p)$ to (X, τ).*
2. *The mapping defined by $g \to g^{-1}\, \forall g \in X$ is continuous from (X, τ) to (X, τ).*

An example of a compact topological group is the unitary group $U(d)$ under the usual operation of multiplication. A compact Lie group [91] is, loosely speaking, a differentiable manifold with a group operation that is smooth with respect to the defined manifold. $U(d)$ is a compact Lie group as a submanifold of $\mathbb{C}^{d^2}$.

D.1.2 Measure Theory

The following is a basic introduction to measure theory. A reference for further reading is [25].

Definition 32. *Algebra, σ-Algebra, Measurable Space*

Let X be a set and $\mathcal{M}$ be a subset of $\mathcal{P}(X)$. $\mathcal{M}$ is called an algebra of sets if

1. $\emptyset \in \mathcal{M}$.
2. $A, B \in \mathcal{M} \Rightarrow A \cup B \in \mathcal{M}$.
3. $A \in \mathcal{M} \Rightarrow X \setminus A \in \mathcal{M}$.

If the second property is extended to countable unions, $\mathcal{M}$ is called a σ-algebra of sets. In this case the ordered pair (X, M) is called a measurable space.

Definition 33. *Measure, Measure Space, Measurable Sets and Probability Measure*

Let (X, M) be a measurable space. A function $\mu : \mathcal{M} \to \mathbb{R} \cup \{\infty\}$ is called a measure if

1. $\mu(\emptyset) = 0$.
2. *For any countable disjoint collection of sets X_i, $\mu(\cup_i X_i) = \sum_i \mu(X_i)$.*
3. $\mu(A) \geq 0 \, \forall A \in \mathcal{M}$.

The triple $(X, \mathcal{M}, \mu)$ is called a measure space and elements of $\mathcal{M}$ are called measurable sets. If in addition to the above conditions $\mu(X) = 1$ then μ is called a probability measure.

An example of a probability measure on a finite set X is the counting measure μ_C defined by

$$\mu_C(A) = \frac{|A|}{|X|} \tag{D.1}$$

for any subset A of X.

Functions that "preserve" the measurable structure between two measurable spaces are of great importance in measure theory. These functions are themselves called "measurable" and are defined as follows:

Definition 34. *Measurable Function*

Let (X, M) and (Y, N) be two measurable spaces. A function $f : X \to Y$ is called measurable if $\forall W \in N$, $f^{-1}(W) \in M$.

For any set X, $\mathcal{P}$(X) is a σ-algebra of subsets of X. Hence, if $\mathcal{S}$ is a subset of $\mathcal{P}(X)$ one can define the Borel algebra of $\mathcal{S}$, $B(\mathcal{S})$, as the smallest σ-algebra containing $\mathcal{S}$. In the case of a topological space (X, τ), the Borel algebra on (X, τ), $B(X, \tau)$ is the smallest σ-algebra containing all of the open sets of τ. A measure defined on $B(X, \tau)$ is called a Borel measure on (X, τ). The following is an important result for Borel measures on compact groups.

Theorem 9. *If (X, τ) is a compact topological group then there exists, up to a constant, a unique Borel measure μ_H on (X, τ), called the bi-invariant Haar measure, satisfying the following conditions*

1. $\mu_H(xE) = \mu_H(E) = \mu_H(Ex) \, \forall x \in X \, \forall E \in B(X, \tau)$.
2. $\mu_H(U) > 0$ *for every non-empty open set* $U \in \tau$.
3. $\mu_H(K) < \infty$ *for every compact set* K.

Since the bi-invariant Haar measure is unique up to a constant and the third property implies $\mu_H(X) < \infty$ there exists a unique bi-invariant Haar probability measure on a compact group.

D.2 Concentration of Measure

Concentration of measure [101, 59, 114, 90] is a phenomenon that can be understood empirically by considering an unbiased coin-tossing experiment consisting of N trials where N is large. Let X be the state space composed of sequences of single trial outcomes. Define the function f from X to $\mathbb{N}$ by $f(x) =$ the number of heads observed. Empirically, $f(x)$ is concentrated around the median value $\frac{N}{2}$ for f. Thus, under the counting measure μ_C on X, $\mu_C(f^{-1}(N-n, N+n))$ is close to 1 even for small $n \in \mathbb{N}$ as N grows large. The following discussion attempts to make these ideas rigorous. For further details of the material here see for instance Ref. [90].

Suppose $(X, d, \mathbb{P})$ is a metric space with Borel probability measure $\mathbb{P}$. To begin, we present a series of useful definitions.

Definition 35. *Diameter*

The diameter of X, diam(X), is defined to be

$$diam(X) = \sup\{d(x, y) \ : x, y \in X\}. \tag{D.2}$$

Definition 36. *ϵ-Neighbourhood*

For $S \subseteq X$ we define

$$N_\epsilon(S) = \{x \in X : \exists y \in S \text{ with } d(x, y) < \epsilon\} \tag{D.3}$$

and call it the ϵ-neighbourhood of S.

Definition 37. *Median*

Let $f : X \to \mathbb{R}$ be a measurable function on X. A median of f, denoted M(f), is defined by the inequalities

$$\mathbb{P}\left[f \leq M(f)\right] \geq \frac{1}{2} \tag{D.4}$$

and

$$\mathbb{P}\left[f \geq M(f)\right] \geq \frac{1}{2}. \tag{D.5}$$

Note that the above inequalities do not imply that a median $M(f)$ satisfies $\mathbb{P}\left[f = M(f)\right] = \frac{1}{2}$ (for instance, consider a constant function). As well, medians may not be unique as can be seen by considering measurable functions whose image consists of two points.

Definition 38. *Modulus of Continuity*

Let $f : X \to \mathbb{R}$ be continuous and $\epsilon > 0$. The modulus of continuity for f given ϵ, denoted $\omega_f(\epsilon)$, is

$$\omega_f(\epsilon) = \sup\{|f(x) - f(y)| : d(x, y) \leq \epsilon\}. \tag{D.6}$$

Definition 39. *K-Lipschitz Functions*

A function $f : X \to \mathbb{R}$ is called K-Lipschitz if $\forall x, y \in X$,

$$|f(x) - f(y)| \leq K\, d(x, y) \tag{D.7}$$

Definition 40. *Concentration Function*

$\forall \epsilon > 0$ the concentration function of X with respect to ϵ is

$$\alpha_X(\epsilon) = 1 - \inf\left\{\mathbb{P}\left[N_\epsilon(S)\right] : S \subseteq X \ \textit{ is Borel measurable and } \ \mathbb{P}(S) \geq \frac{1}{2}\right\}. \tag{D.8}$$

Note that the concentration functions decrease as ϵ grows. An equivalent expression for the concentration function is,

$$\alpha_X(\epsilon) = \sup\left\{\mathbb{P}\left[X - N_\epsilon(S)\right] : \text{ S} \subseteq X \text{ is Borel measurable and } \ \mathbb{P}(S) \geq \frac{1}{2}\right\}. \tag{D.9}$$

The concept of concentration of measure is easiest seen using the concentration function as defined above. When the concentration functions are very small in ϵ then the space is concentrated in that ϵ neighbourhoods of sets with measure at least $\frac{1}{2}$ have measure very

close to 1. There is an equivalent characterization of this property in terms of continuous functions. Note that for f continuous with modulus of continuity $\omega_f(\epsilon)$,

$$\begin{aligned}
\mathbb{P}\left[f \geq M(f) + \omega_f(\epsilon)\right] &= \mathbb{P}\left[X - [f \leq M(f) + \omega_f(\epsilon)]\right] \\
&= 1 - \mathbb{P}\left[f \leq M(f) + \omega_f(\epsilon)\right] \\
&\leq 1 - \mathbb{P}\left[N_\epsilon\left([f \leq M(f) + \omega_f(\epsilon)]\right)\right] \\
&\leq \alpha_X(\epsilon). \qquad \text{(D.10)}
\end{aligned}$$

Hence,

$$\sup\{\mathbb{P}\left[f \geq M(f) + \omega_f(\epsilon)\right] \,:\, f \text{ is continuous}\} \leq \alpha_X(\epsilon). \qquad \text{(D.11)}$$

From the above inequality, the idea of concentration of measure is equivalent to the clustering of any continuous function about a median of the function. We now define Levy and normal Levy families which are important due to concentration of measure often occurring in an asymptotic sense for a wide variety of spaces.

Definition 41. *Levy Family*

$\forall i \in \mathbb{N}$ let $\mathcal{F} = \{((X_i, d_i), \mathbb{P}_i)\}$ be a family of metric spaces with Borel probability measures $\mathbb{P}_i$ and diameters $diam(X_i)$. $\mathcal{F}$ is called Levy if $\forall \epsilon > 0$,

$$\lim_{i\to\infty} \alpha_{X_i}\left(\epsilon\, diam(X_i)\right) = 0. \qquad \text{(D.12)}$$

The factor of $\text{diam}(X_i)$ in the argument of the concentration function is required because we are fixing ϵ independent of i and then taking the limit as $i \to \infty$. If the factor of $\text{diam}(X_i)$ is left out of the definition then for a sequence of spaces with diameters that converge to 0 quickly as $i \to \infty$, it would be the case that these sequences trivially satisfy the requirement for being concentrated.

Note that from page 56 of [90], $\mathcal{F}$ is Levy if for $\epsilon > 0$ and any sequence of Borel sets $S_i \subseteq X_i$ such that $\liminf_{i\to\infty} \mathbb{P}_i(S_i) > 0$,

$$\lim_{i\to\infty} \mu_i(N_{\epsilon diam(X_i)}(S_i)) = 1. \qquad \text{(D.13)}$$

Definition 42. *Normal Levy Family*

$\forall i \in \mathbb{N}$ *let* $\mathcal{F} = \{((X_i, d_i), \mathbb{P}_i)\}$ *be a family of metric spaces with Borel probability measures* $\mathbb{P}_i$ *and diameters* $diam(X_i)$. $\mathcal{F}$ *is called a normal Levy family if there exist constants* A, B *such that* $\forall i$ *and* $\epsilon > 0$

$$\alpha_{X_i}(\epsilon) \leq Ae^{-B\epsilon^2 i}. \tag{D.14}$$

Note that it may not be true that a normal Levy family is a Levy family. To see this, if $\mathcal{F}$ is a normal Levy family then for every i and $\epsilon > 0$,

$$\alpha_{X_i}(\epsilon\, diam(X_i)) \leq Ae^{-B\epsilon^2 (diam(X_i))^2 i} \tag{D.15}$$

which may not converge to 0 as $i \to \infty$ because $(diam\,(X_i))^2\, i$ may not converge to ∞. If say

$$\lim_{i\to\infty} (diam\,(X_i))^2 > 0 \tag{D.16}$$

then the normal Levy family would be a Levy family.

The following lemma is required for the main result regarding normal Levy families.

Lemma 2. *Let* $\epsilon > 0$ *and* f *be a continuous function on* $(X, d, \mathbb{P})$ *with modulus of continuity* $\omega_f(\epsilon)$. *Then,*

$$\mathbb{P}\left[|f - M(f)| < \omega_f\left(\epsilon\right)\right] \geq 1 - 2\alpha_X(\epsilon). \tag{D.17}$$

Proof. First note that,

$$\mathbb{P}\left[|f - M(f)| < \omega_f\left(\epsilon\right)\right] = \mathbb{P}\left[\left[f \leq M(f) + \omega_f\left(\epsilon\right)\right] \cap \left[f \geq M(f) - \omega_f\left(\epsilon\right)\right]\right]. \tag{D.18}$$

By the definition of modulus of continuity,

$$N_\epsilon\left(\left[f \leq M(f)\right]\right) \subseteq \left[f \leq M(f) + \omega_f\left(\epsilon\right)\right] \tag{D.19}$$

and,

$$N_\epsilon\left([f \geq M(f)]\right) \subseteq \left[f \geq M(f) - \omega_f\left(\epsilon\right)\right]. \tag{D.20}$$

Hence,

$$\mathbb{P}\left[[f \leq M(f) + \omega_f\left(\epsilon\right)] \cap [f \geq M(f) - \omega_f\left(\epsilon\right)]\right] \geq \mathbb{P}\left[N_\epsilon\left([f \leq M(f)]\right) \cap N_\epsilon\left([f \geq M(f)]\right)\right]. \tag{D.21}$$

Now for any measurable sets A and B,

$$\begin{aligned} \mathbb{P}\left(A \cap B\right) &= \mathbb{P}\left(A^C \cup B^C\right)^C \\ &= 1 - \mathbb{P}\left(A^C \cup B^C\right) \\ &\geq 1 - \mathbb{P}\left(A^C\right) - \mathbb{P}\left(B^C\right). \end{aligned} \tag{D.22}$$

Since each of $[f \leq M(f)]$ and $[f \geq M(f)]$ have measure bounded below by $\frac{1}{2}$, it follows by the definition of the concentration function that

$$\mathbb{P}\left[N_\epsilon\left([f \leq M(f)]\right)\right] \leq \alpha_X(\epsilon) \tag{D.23}$$

and,

$$\mathbb{P}\left[N_\epsilon\left([f \geq M(f)]\right)\right] \leq \alpha_X(\epsilon). \tag{D.24}$$

Hence in total,

$$\mathbb{P}\left[[f \leq M(f) + \omega_f\left(\epsilon\right)] \cap [f \geq M(f) - \omega_f\left(\epsilon\right)]\right] \geq 1 - 2\alpha_X(\epsilon). \tag{D.25}$$

which is the desired result.

□

The following theorem is the main result and is a direct consequence of the above lemma.

Theorem 10. *Let $\mathcal{F}$ be a normal Levy family. For each i, let f_i be a continuous function on $(X_i, d_i, \mathbb{P}_i)$ with median $M(f_i)$ and modulus of continuity $\omega_{f_i}(\epsilon)$ for each $\epsilon > 0$. Then $\forall \epsilon > 0$,*

$$\begin{aligned}\mathbb{P}\left[|f_i - M(f_i)| < \omega_{f_i}(\epsilon)\right] &\geq 1 - 2\alpha_{X_i}(\epsilon) \\ &\geq 1 - 2Ae^{-B\epsilon^2 i}. \end{aligned} \tag{D.26}$$

Next we look at various examples of normal Levy families.

D.3 Examples of Concentration of Measure

Example 1. *The standard example of a normal Levy family is that of unit spheres in* $(\mathbb{R}^n, \| \|_2)$ *with the geodesic metric. The 2-norm,* $\| \|_2$*, on* $\mathbb{R}^n$*, is defined through the Euclidean inner product on* $\mathbb{R}^n$*. Taking* $x = (x_1, ..., x_n) \in \mathbb{R}^n$*,*

$$||x||_2 := \sqrt{< x, x >} = \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{D.27}$$

As usual,

$$\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : ||x|| = 1\}. \tag{D.28}$$

Let d *be the geodesic (Riemannian) metric on* $\mathbb{S}^{n-1}$*,*

$$d(x, y) = arccos < x, y > \tag{D.29}$$

which is the angle between x *and* y *in* $\mathbb{R}^n$*. As well, let* $\mathbb{P}$ *be the unique Borel (Haar) measure on* $\mathbb{S}^{n-1}$ *generated by the topology induced by* d*. Suppose f:* $\mathbb{S}^{n-1} \to \mathbb{R}$ *is continuous and let M(f) be the median of f. Then,*

$$\mathbb{P}\left[|f - M(f)| \leq \epsilon\right] \geq 1 - \sqrt{\frac{\pi}{2}} e^{-\epsilon^2 \frac{(n-2)}{2}}. \tag{D.30}$$

Thus the set $\{(\mathbb{S}^{n+1} \subset \mathbb{R}^{n+2}, d_{n+1}, \mathbb{P}_{n+1})\}$ *indexed by* $n \in \mathbb{N}$ *is a Levy family with* $A = \sqrt{\frac{\pi}{8}}$ *and* $B = \frac{1}{2}$*.*

Example 2. *Next, we look at the unitary group U(n). $\forall n \in \mathbb{N}$, $U(n)$ is a Lie group. In addition, $U(n)$ is compact and so can be equipped with a unique bi-invariant (ie. left and right invariant) metric. Let d_n be this bi-invariant metric on $U(n)$ (which is induced by the trace inner product),*

$$d(U,V) = \sqrt{tr((U-V)^{\dagger}(U-V))} \tag{D.31}$$

Denote the Haar measure on $U(n)$ by $\mathbb{P}_n$. The following theorem is proved in [101].

Theorem 11. *The family $\{(U(n), d_n), \mu_n\}$ is a normal Levy family with constants $A = \sqrt{\frac{\pi}{8}}$ and $B = \frac{1}{8}$.*

Example 3. *This example also deals with $\mathbb{S}^n$ defined above, however the metric defined on it is the Euclidean metric $\| \ \|_2$. In this case, by Appendix V of [101], we have for an η-Lipschitz function $f : \mathbb{S}^n \to \mathbb{R}$ (with respect to $\| \ \|_2$) and the Haar measure $\mathbb{P}$ of Example 1,*

$$\mathbb{P}\left[f < \int f d\mu - \epsilon\right] \leq 2e^{\frac{-C\epsilon^2(n+1)}{\eta^2}} \tag{D.32}$$

and

$$\mathbb{P}\left[f > \int f d\mu + \epsilon\right] \leq 2e^{\frac{-C\epsilon^2(n+1)}{\eta^2}} \tag{D.33}$$

where $C = \frac{1}{9\pi^3 ln2}$ and $\int f d\mu$ is the integral of f with respect to the Haar measure. This implies,

$$\mathbb{P}\left[\left|f - \int f d\mu\right| < \epsilon\right] \geq 1 - 4e^{\frac{-C\epsilon^2(n+1)}{\eta^2}}. \tag{D.34}$$

Additionally, a relationship between the measure of $\left[\left|f - \int f d\mu\right| < \epsilon\right]$ and $[|f - M(f)| < \epsilon,]$ is given which results in analogous inequalities,

$$\mathbb{P}\left[f < M(f) - \epsilon\right] \leq e^{\frac{-D\epsilon^2(n-1)}{\eta^2}}, \tag{D.35}$$

$$\mathbb{P}\left[f > M(f) + \epsilon\right] \leq e^{\frac{-D\epsilon^2(n-1)}{\eta^2}}, \tag{D.36}$$

and

$$\mathbb{P}\left[|f - M(f)| < \epsilon\right] \geq 1 - 2e^{\frac{-D\epsilon^2(n-1)}{\eta^2}} \tag{D.37}$$

where $D = \frac{1}{2\pi^2 ln2}$.

Other examples of Normal Levy families are the permutation groups and Hamming cubes $\{0,1\}^n$ of all binary strings of length n. Both are equipped with the normalized Hamming distance and the normalized counting measure. Further examples can be found in in [114, 101, 90].

Appendix E

Symplectic Representation and Decomposing Clifford Group Elements

In this section we provide a brief introduction to the symplectic representation of the Clifford group, as well as how to decompose Clifford group elements into a sequence of generators for the Clifford group.

E.1 Symplectic Representation of the Clifford Group

The general idea behind the symplectic representation is to first associate Pauli matrices and matrix multiplication to binary vector spaces with addition. Representing Pauli matrices by binary vector spaces is common in various areas of quantum information, for instance in quantum error correction [57] where one can represent a set of generators for a stabilizer code using check matrices. From the association of Pauli matrices to binary spaces, one can represent Clifford operations via linear operations (specifically symplectic operations) on these spaces.

We follow the set-up and notation used in [41] and define

$$\begin{aligned}\sigma_{00} &= \mathbb{1}, \\ \sigma_{01} &= X, \\ \sigma_{10} &= Z, \\ \sigma_{11} &= Y,\end{aligned} \tag{E.1}$$

and

$$\begin{aligned}\tau_{00} &= \sigma_{00} = \mathbb{1}, \\ \tau_{01} &= \sigma_{01} = X, \\ \tau_{10} &= \sigma_{10} = Z, \\ \tau_{11} &= i\sigma_{11} = iY.\end{aligned} \tag{E.2}$$

Hence if $a \in \mathbb{Z}_2^2$ we associate τ_a with a Pauli operator (possibly multiplied by i) and note that if $a = (a_1, a_2)$ then $\tau_a = \sigma_{10}^{a_1}\sigma_{01}^{a_2} = Z^{a_1}X^{a_2}$. This definition can easily be extended to n qubits by defining for $v, w \in \mathbb{Z}_2^n$ and $a := (v, w) \in \mathbb{Z}_2^{2n}$,

$$\begin{aligned}\sigma_a &= \sigma_{v_1,w_1} \otimes ... \otimes \sigma_{v_n,w_n} \\ \tau_a &= \tau_{v_1,w_1} \otimes ... \otimes \tau_{v_n,w_n}.\end{aligned} \tag{E.3}$$

An arbitrary element of the Pauli group on n qubits, $\mathcal{P}_n$, is thus given by $i^\delta(-1)^\epsilon\tau_a$ where δ, $\epsilon \in \mathbb{Z}_2$ and $a \in \mathbb{Z}_2^{2n}$.

Lemma 1 of [41] gives that if $i^{\delta_1}(-1)^{\epsilon_1}\tau_{a_1}$ and $i^{\delta_2}(-1)^{\epsilon_2}\tau_{a_2}$ are elements of $\mathcal{P}_n$ then,

$$\left[i^{\delta_1}(-1)^{\epsilon_1}\tau_{a_1}\right]\left[i^{\delta_2}(-1)^{\epsilon_2}\tau_{a_2}\right] = i^{\delta_{12}}(-1)^{\epsilon_{12}}\tau_{a_{12}} \tag{E.4}$$

where

$$\begin{aligned}
\delta_{12} &= \delta_1 + \delta_2, \\
\epsilon_{12} &= \epsilon_1 + \epsilon_2 + \delta_1\delta_2 + a_2^T U a_1, \\
a_{12} &= a_1 + a_2, \\
U &= \begin{pmatrix} 0_n & \mathbb{1}_n \\ 0_n & 0_n \end{pmatrix}.
\end{aligned} \tag{E.5}$$

It is easy to see how all of these terms arise (note that $i^{\delta_1} i^{\delta_2} = i^{\delta_1+\delta_2}(-1)^{\delta_1\delta_2}$ in binary arithmetic) except perhaps the term $a_2^T U a_1$ in ϵ_{12}. This term adds (modulo 2) how many negative signs appear in the n bit-wise multiplication terms. To see this suppose $a_1 = (v_1, w_1)$ and $a_2 = (v_2, w_2)$ so that $a_2^T U a_1$ counts (modulo 2) the number of positions k where $v_{2k} = w_{1k} = 1$. Hence since the single-qubit commutation relations imply

$$\begin{aligned}
\tau_{v_{1k}w_{1k}}\tau_{v_{2k}w_{2k}} &= \sigma_z^{v_{1k}}\sigma_x^{w_{1k}}\sigma_z^{v_{2k}}\sigma_x^{w_{2k}} \\
&= (-1)^{w_{1k}v_{2k}}\sigma_z^{v_{1k}+v_{2k}}\sigma_x^{w_{1k}+w_{2k}} \\
&= (-1)^{w_{1k}v_{2k}}\tau_{v_{1k}+v_{2k},w_{1k}+w_{2k}}
\end{aligned} \tag{E.6}$$

the total number of places where a negative sign occurs is given by $a_2^T U a_1$.

Every Clifford operation $\mathcal{Q}(X) = QXQ^\dagger$ is uniquely determined (up to phase) by its action on a generating set for $\mathcal{P}_n$. One convenient choice of generating set to analyze is the $2n$ elements of $\mathcal{P}_n$ consisting of single-qubit σ_x and σ_z operators. In the binary picture for $\mathcal{P}_n$ this corresponds to knowing the images of the $2n$ standard basis vectors e_k of $\mathbb{Z}_2^{2n}$ under $\mathcal{Q}$. Let us denote the image of each e_k under $\mathcal{Q}$ by $i^{d_k}(-1)^{h_k}\tau_{c_k}$ and represent this information via three matrices: a $2n$ by $2n$ matrix C whose $2n$ columns are the c_k vectors, a $2n$ by 1 matrix d whose entries are the d_k and a $2n$ by 1 matrix h whose entries are the h_k. Since the images of each e_k under $\mathcal{Q}$ are Hermitian, $d_k = c_k^T U c_k$ which implies $d = \mathrm{diag}(C^T U C)$.

Now since C, d and h uniquely define a Clifford operation (up to phase) we can find the image of any element $i^{\delta_1}(-1)^{\epsilon_1}\tau_{b_1} \in \mathcal{P}_n$ under $\mathcal{Q}$ by first noting that $i^{\delta_1}(-1)^{\epsilon_1}\tau_{b_1}$ is, up to factors of ± 1 and $\pm i$, equal to the product of all elements of the standard basis that are associated with a value of "1" in the length $2n$ vector b_1. Hence the image of $i^{\delta_1}(-1)^{\epsilon_1}\tau_{b_1}$ under $\mathcal{Q}$ is, up to multiplicative factors, equal to the product of all $i^{d_k}(-1)^{h_k}\tau_{c_k}$ for which $b_{1k} = 1$. As noted in [41] this gives,

$$
\begin{aligned}
b_2 &= Cb_1, \\
\delta_2 &= \delta_1 + d^T b_1, \\
\epsilon_2 &= \epsilon_1 + h^T b_1 + b_1^T \text{Lower}\left(C^T U C + d d^T\right) b_1 + \delta_1 d^T b_1
\end{aligned} \tag{E.7}
$$

where the operation Lower(X) corresponds to the (strictly) lower triangular part of X.

Next, it is discussed in [41] that $\mathcal{Q}$ is a Clifford operation if and only if the matrix C is symplectic and $d = \text{diag}\left(C^T U C\right)$ (which states that Hermiticity is preserved when $\mathcal{Q}$ operates on Hermitian elements). By "symplectic" it is meant that C satisfies

$$
C^T P C = P \tag{E.8}
$$

where

$$
P = U + U^T = \begin{pmatrix} 0_n & \mathbb{1}_n \\ \mathbb{1}_n & 0_n \end{pmatrix}. \tag{E.9}
$$

Note that

$$
P = \begin{pmatrix} 0_n & \mathbb{1}_n \\ -\mathbb{1}_n & 0_n \end{pmatrix} \tag{E.10}
$$

is the standard skew-symmetric matrix symplecticity is usually defined with respect to. The commutation relations of Pauli operators can be phrased in terms of the binary formalism via,

$$
\tau_a \tau_b = (-1)^{b^T P a} \tau_b \tau_a \tag{E.11}
$$

where $a \odot b := b^T P a$ represents the "symplectic inner product" on $\mathbb{Z}_2^{2n}$. To see why every Clifford operation must be represented by a symplectic matrix C note that Clifford operations preserve commutation relations. Hence if a and b represent commuting Pauli operators (ie. $b^T P a = 0$) then the images of these operators under a Clifford operation must satisfy $b^T C^T P C a = 0$. Thus $C^T P C$ is a matrix which exactly describes the commutation relation in Eq. (E.11) and so it must be that $C^T P C = P$, ie. C is symplectic. The fact that

every symplectic matrix corresponds to a Clifford operation can be shown by decomposing C into a product of matrices (or more importantly for our purposes, generators) which are known to represent Clifford operations.

Before discussing how to decompose a symplectic matrix C into a sequence of generators from the Clifford group we note the form that C and h take for various cases that will be of relevance. We only list the results as they are straightforward to verify (or see [41]).

1. If $\mathcal{Q}$ is a Pauli operator, ie. $Q = \tau_a$ for $a \in \mathbb{Z}_2^{2n}$, then,

$$C = \mathbb{1}_{2n}, \quad h = Pa. \tag{E.12}$$

2. The Hadamard operation on a single qubit is represented by

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad h = 0. \tag{E.13}$$

3. The single qubit phase gate S has representation

$$C = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad h = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{E.14}$$

4. Let Π represent a permutation matrix of the n qubits. Then,

$$C = \begin{pmatrix} \Pi & 0 \\ 0 & \Pi \end{pmatrix}, \quad h = 0. \tag{E.15}$$

5. The $CNOT$ operator (acting on two qubits) is represented by,

$$C = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad h = 0. \tag{E.16}$$

6. Let R represent any invertible matrix on $\mathbb{Z}_2^n$ (ie. $R \in \mathbb{Z}_2^{n \times n}$). The linear mapping induced by the invertible linear transformation R on the index space ($|x\rangle \to |Rx\rangle$) is a Clifford operation and is represented by

$$C = \begin{pmatrix} R^{-T} & 0 \\ 0 & R \end{pmatrix}, \quad h = 0. \tag{E.17}$$

Note that qubit permutations and $CNOT$ both have C matrices of this form and it is easy to verify that $|x\rangle \to |\Pi x\rangle$ and $|x\rangle \to |Nx\rangle$ respectively, where

$$N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{E.18}$$

In fact *any* invertible matrix R on index space can be decomposed into a sequence of two-level operations composed of $CNOT$ and two-qubit permutation matrices. This is because the elementary row operations that are used in Gaussian elimination on elements of $\mathbb{Z}_2^{n\times n}$ are exactly given by the C matrices for $CNOT$ and two-qubit permutation matrices. Since R is assumed to be invertible the product of all of the operations used to bring R into RREF is equal to R and hence R can be decomposed into $CNOT$ operations and two-qubit permutation matrices.

7. Let τ_a represent a Pauli operation and consider the Hermitian version of τ_a, $\tau_{\bar{a}} = i^{a^T U a}\tau_a$. The rotation operator

$$e^{i\frac{\pi}{4}\tau_{\bar{a}}} = \frac{1}{\sqrt{2}}\left(\mathbb{1} + i\tau_{\bar{a}}\right) \tag{E.19}$$

is represented by

$$C = \mathbb{1} + aa^T P, \quad h = C^T U a. \tag{E.20}$$

8. Here we look at Clifford operations that act non-trivially only on a subset $\alpha \subseteq \{1, ..., n\}$ of the n qubit system. In this case we have a symplectic matrix on the rows and columns with indices in $\alpha \cup (\alpha + n)$ that is embedded within an identity matrix. More precisely, $C_{k,k} = 1$ if $k \notin \alpha \cup (\alpha + n)$ and $C_{k,l} = 0$ if both $k \neq l$ and either k or l are not in $\alpha \cup (\alpha + n)$. Lastly, h is such that $h_k = 0$ if $k \notin \alpha \cup (\alpha + n)$.

E.2 Decomposing Clifford Group Elements

In this section we give a brief overview of the method presented in [41] for decomposing Clifford operations into a sequence of generators. We have that every Clifford element $\mathcal{Q}$ is represented up to phase by a symplectic matrix $C \in \mathbb{Z}_2^{2n\times 2n}$ and a vector $h \in \mathbb{Z}_2^{2n\times 1}$. It is noted in [41] that the main goal is to decompose C into generators for the negative signs represented by h can be introduced via multiplication by single-qubit Pauli operators (which are obviously in the Clifford group). We discuss this point in more detail afterwards. The main theorem for the decomposition of Clifford elements is the following [41]:

If $C \in \mathbb{Z}_2^{2n\times 2n}$ is symplectic then:

$$C = \begin{pmatrix} T_1^{-T} & 0 \\ 0 & T_1 \end{pmatrix} \begin{pmatrix} \mathbb{1}_{n-r} & V_1 & Z_3 + V_1 V_2^T & V_2 + V_1 Z_2 \\ 0 & Z_1 & V_1^T + Z_1 V_2^T & \mathbb{1}_r + Z_1 Z_2 \\ 0 & 0 & \mathbb{1}_{n-r} & 0 \\ 0 & \mathbb{1}_r & V_2^T & Z_2 \end{pmatrix} \begin{pmatrix} T_2^{-T} & 0 \\ 0 & T_2 \end{pmatrix} \quad \text{(E.21)}$$

where the middle matrix of the right-hand side is equal to

$$\begin{pmatrix} \mathbb{1}_{n-r} & 0 & Z_3 & V_1 \\ 0 & \mathbb{1}_r & V_1^T & Z_1 \\ 0 & 0 & \mathbb{1}_{n-r} & 0 \\ 0 & 0 & 0 & \mathbb{1}_r \end{pmatrix} \begin{pmatrix} \mathbb{1}_{n-r} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1}_r \\ 0 & 0 & \mathbb{1}_{n-r} & 0 \\ 0 & \mathbb{1}_r & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbb{1}_{n-r} & 0 & 0 & V_2 \\ 0 & \mathbb{1}_r & V_2^T & Z_2 \\ 0 & 0 & \mathbb{1}_{n-r} & 0 \\ 0 & 0 & 0 & \mathbb{1}_r \end{pmatrix} \quad \text{(E.22)}$$

and

- r is defined in the proof below,
- T_1 and T_2 are invertible elements of $\mathbb{Z}_2^{n\times n}$,
- Z_1 and Z_2 are symmetric elements of $\mathbb{Z}_2^{r\times r}$
- Z_3 is a symmetric element of $\mathbb{Z}_2^{(n-r)\times(n-r)}$,

- V_1 and V_2 are elements of $\mathbb{Z}_2^{(n-r)\times r}$,
- the blocks consisting of zero's have appropriate size.

The proof of this theorem is as follows [41]: First, write C as a block matrix of four elements of $\mathbb{Z}_2^{n\times n}$,

$$C = \begin{pmatrix} E' & F' \\ G' & H' \end{pmatrix}. \tag{E.23}$$

Next, find invertible R_1 and R_2 in $\mathbb{Z}_2^{n\times n}$ that satisfy

$$R_1^{-1} G' R_2 = \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1}_r \end{pmatrix}. \tag{E.24}$$

where $r = \text{rank}\,(G')$. This can be done by looking at the kernel and image of G', denoted $\ker(G')$ and $\text{im}\,(G')$ respectively. $\ker(G')$ corresponds to the linear subspace of $\mathbb{Z}_2^n$ which gets mapped to 0 under G' and $\text{im}\,(G')$ corresponds to the linear subspace of $\mathbb{Z}_2^n$ for which each element of $\text{im}\,(G')$ is mapped to by some element of $\mathbb{Z}_2^n$ under G'. Hence $r = \dim(\text{im}\,(G'))$ and $\dim(\text{im}\,(G')) + \dim(\ker(G')) = n$ implies $\dim(\ker(G')) = n - r$.

Thus, let the first $n-r$ columns of R_2 be a basis for $\ker(G')$ which implies the first $n-r$ columns of $G'R_2$ are zero. This is done by bringing G' into RREF and constructing a basis from the columns that don't contain a leading 1. Next, choose the remaining r columns of R_2 so that R_2 is invertible (and thus the image of these columns under G' will form a basis for $\text{im}\,(G')$). This can be done in the following manner:

Since $\dim(\ker(G')) = n - r$, $\ker(G')$ is an $n-r$-dimensional vector space (so contains 2^{n-r} vectors) that has basis given by the first $n-r$ columns of R_2, which we denote $\{b_1, ..., b_{n-r}\}$. Hence there are $2^n - 2^{n-r}$ vectors in $\mathbb{Z}_2^n$ such that when any one is added to $\{b_1, ..., b_{n-r}\}$ the resulting set is linearly independent. Hence if we choose elements of $\mathbb{Z}_2^n$ uniformly at random then with probability $p_{n-r} = \frac{2^n - 2^{n-r}}{2^n} = 1 - \frac{1}{2^r}$ such an element will be linearly independent with $\{b_1, ..., b_{n-r}\}$. Assuming we have found such a vector we now have the linearly independent set $\{b_1, ..., b_{n-r+1}\}$ and we repeat this process $r-1$ more times to obtain a basis for the kernel. The probability of succeeding at each and every one of the r steps is

$$\left(1 - \frac{1}{2^r}\right)\left(1 - \frac{1}{2^{r-1}}\right) ... \left(1 - \frac{1}{2}\right) \tag{E.25}$$

This probability can be increased by repeating the process k times at each of the r steps. In this case, at say the first step, a probability of (at least one) success is $1 - \frac{1}{2^{kr}}$ (since the probability of failing in every trial is $\left(1 - \left(1 - \frac{1}{2^r}\right)\right)^k = \frac{1}{2^{kr}}$). More generally, at the j'th step the probability of (at least one) success is $1 - \frac{1}{2^{k(r-j+1)}}$. Thus with k trials at each step, the total probability of succeeding over the r steps is,

$$\left(1 - \frac{1}{2^{kr}}\right)\left(1 - \frac{1}{2^{k(r-1)}}\right)...\left(1 - \frac{1}{2^k}\right) \tag{E.26}$$

which has lower bound independent of r given by $\left(1 - \frac{1}{2^k}\right)^n$.

Next, note that we want R_1 to satisfy $G'R_2 = R_1 \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1}_r \end{pmatrix}$. Hence, choose the last r columns of R_1 to be equal to the last r columns of $G'R_2$. As noted above these vectors form a basis for $\mathrm{im}\,(G')$. The first $n - r$ columns of R_1 are now chosen so that R_1 is invertible. The procedure for doing this is the same as described above for extending R_2 to an invertible matrix.

R_1 and R_2 have been constructed in this manner so that we can write

$$\begin{pmatrix} R_1^T & 0 \\ 0 & R_1^{-1} \end{pmatrix} C \begin{pmatrix} R_2 & 0 \\ 0 & R_2^{-T} \end{pmatrix} = \begin{pmatrix} E_{11} & E_{12} & F_{11} & F_{12} \\ E_{21} & E_{22} & F_{21} & F_{22} \\ 0 & 0 & H_{11} & H_{12} \\ 0 & \mathbb{1}_r & H_{21} & H_{22} \end{pmatrix} \tag{E.27}$$

where E_{11} is $(n-r) \times (n-r)$, E_{12} is $(n-r) \times r$, E_{21} is $r \times (n-r)$ and E_{22} is $r \times r$ and similarly for the F and H matrices. In order to obtain this decompositon we need to explicitly find R_1^{-1} and R_2^{-1}, however this follows in a straightforward from performing Gaussian elimination on invertible matrices. The inverse is constructed from the elementary operations corresponding to the row operations performed to bring the matrix to RREF (which in this case is the identity). As noted, the row operations performed are switching two rows and zeroing one row with respect to another, which correspond to two-qubit permutations and CNOT gates respectively.

From Example 6 of the various symplectic representations of particular Clifford operations we see that each of $\begin{pmatrix} R_1^T & 0 \\ 0 & R_1^{-1} \end{pmatrix}$ and $\begin{pmatrix} R_2 & 0 \\ 0 & R_2^{-T} \end{pmatrix}$ are symplectic. Hence since C is symplectic, the RHS of the above equation is symplectic. Thus the construction of R_1 and R_2 above allow us to write an equation of the form Eq. (E.27) such that the RHS

satisfies the important properties of being symplectic with the lower left corner equal to $\begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1}_r \end{pmatrix}$. These properties imply a list of nine relationships that are contained in Eq.'s (8)-(16) in [41]. One of the main results of these relationships is

- $H_{11} = E_{11}^{-T}$.

Therefore if R_2 is replaced with $R_3 := R_2 \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & \mathbb{1}_r \end{pmatrix}$ an equation analogous to Eq. (E.27) holds (symplecticity still holds as well) where the E_{ij}, F_{ij} and H_{ij} are updated versions of those in Eq. (E.27), and both of H_{11} and E_{11} can be replaced by $\mathbb{1}_{n-r}$, ie.

$$\begin{aligned}
&\begin{pmatrix} R_1^T & 0 \\ 0 & R_1^{-1} \end{pmatrix} C \begin{pmatrix} R_3 & 0 \\ 0 & R_3^{-T} \end{pmatrix} \\
&= \begin{pmatrix} R_1^T & 0 \\ 0 & R_1^{-1} \end{pmatrix} C \begin{pmatrix} R_2 & 0 \\ 0 & R_2^{-T} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & \mathbb{1}_r \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & \mathbb{1}_r \end{pmatrix}^{-T} \end{pmatrix} \\
&= \begin{pmatrix} E_{11} & E_{12} & F_{11} & F_{12} \\ E_{21} & E_{22} & F_{21} & F_{22} \\ 0 & 0 & H_{11} & H_{12} \\ 0 & \mathbb{1}_r & H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & \mathbb{1}_r \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} E_{11}^{-1} & 0 \\ 0 & \mathbb{1}_r \end{pmatrix}^{-T} \end{pmatrix} \\
&:= \begin{pmatrix} \mathbb{1}_{n-r} & E_{12} & F_{11} & F_{12} \\ E_{21} & E_{22} & F_{21} & F_{22} \\ 0 & 0 & \mathbb{1}_{n-r} & H_{12} \\ 0 & \mathbb{1}_r & H_{21} & H_{22} \end{pmatrix}.
\end{aligned} \tag{E.28}$$

Eq. (E.28) now implies the following:

- $H_{12} = 0$,
- E_{22} and H_{22} are symmetric,
- $F_{21} = E_{12}^T + E_{22}^T H_{21}$ and $F_{22} = \mathbb{1}_r + E_{22} H_{22}$.

Hence if we define

- $T_1 := R_1$,
- $T_2 = R_3^T$,
- $V_1 = E_{12}$,
- $V_2 = H_{21}^T$,
- $Z_1 = E_{22}$,
- $Z_2 = H_{22}$,
- $Z_3 = F_{11} + V_1 V_2^T$,

then Eq. (E.21) is verified with the associated properties of submatrices listed below the equation and Eq. (E.22) follows by just writing the middle matrix of Eq. (E.21) as a product of three matrices.

We now have C in terms of a product of five matrices each of which is a symplectic element of $\mathbb{Z}_2^{2n\times 2n}$. From Example 6 above, clearly the first and fifth matrices correspond to invertible linear index space transformations. Hence, as noted in the discussion of Example 6, the $n \times n$ sub-matrices T_1 and T_2 can be decomposed into a product of $CNOT$ gates and two-qubit permutation matrices by reducing them to RREF (which in this case is the identity since these matrices are invertible). More precisely, these are the elementary matrices used in Gaussian elimination on matrices over $\mathbb{Z}_2$. Next, from combining Example 2 with Example 8, the third matrix in Eq. (E.22) corresponds to Hadamard matrices on the last r qubits.

Lastly, we have to deal with the second and fourth matrices. Note that these matrices are of the form $\begin{pmatrix} \mathbb{1}_n & Z \\ 0 & \mathbb{1}_n \end{pmatrix}$ with Z a symmetric element of $\mathbb{Z}_2^{n\times n}$. The set of all such matrices forms a commutative subgroup of $\mathbb{Z}_2^{2n\times 2n}$ (which is expected from the commutativity of one and two qubit $\frac{\pi}{2}$ rotations about Z and $Z \otimes Z$), which we denote by T_{2n}. We show how to write any element R of T_{2n} as a product of two-qubit Clifford elements. The main idea is to look at the diagonal and off-diagonal elements of Z separately. By symmetry we need only look at the diagonal and upper off-diagonal elements of Z. First we deal with the off-diagonal elements.

Suppose $Z_{k,l} \neq 0$ with $k < l$ (so $Z_{l,k} = 1$ as well). Let $B \in T_{2n}$ be such that

$$B = \begin{pmatrix} \mathbb{1}_n & V \\ 0 & \mathbb{1}_n \end{pmatrix} \tag{E.29}$$

where V is non-zero only at its (k,l) and (l,k) entries. We can see from Example 7 listed above that if we set $a \in \mathbb{Z}_2^{2n\times 1}$ to be such that $a_k = a_l = 1$ and $a_j = 0$ for all other j then the two-qubit Clifford operation $e^{i\frac{\pi}{4}\tau_{\bar{a}}}$ has symplectic matrix $C = \mathbb{1}_{2n} + aa^T P$ equal to B except that the $n \times n$ upper right corner of C has a value of 1 at both its (k,k) and (l,l) entry. We can remove these unwanted values by applying the single-qubit rotations $e^{i\frac{\pi}{4}\tau_{\bar{b}}}$ and $e^{i\frac{\pi}{4}\tau_{\bar{c}}}$ where b and c are length $2n$ vectors that are non-zero only at k and l. We leave further discussion of this point until we analyze the diagonal elements of Z. Note that since k and l are no larger than n, a is non-zero in only its first n elements which implies the two non-identity factors in τ_a are both σ_Z operators and so $\tau_{\bar{a}} = \tau_a$ as τ_a is Hermitian. Now multiplying all such $B \in T_{2n}$ corresponding to non-zero upper off-diagonal elements exactly produces a matrix $R' = \begin{pmatrix} \mathbb{1}_n & V' \\ 0 & \mathbb{1}_n \end{pmatrix}$ that is equal to R except perhaps the diagonal elements of V' and V may not be equal. We now turn attention to these diagonal elements.

Let us fix $k \in n$ and first suppose $V_{k,k} = 0$. If

$$|\{l \in \{1,...,n\} \,:\, l > k \text{ and } a_{k,l} = 1\}| + |\{l \in \{1,...,n\} \,:\, l < k \text{ and } a_{l,k} = 1\}| = 0 \,(\text{mod} 2) \tag{E.30}$$

(ie. the number of $l > k$ with $a_{k,l} = 1$ plus the number of $l < k$ with $a_{l,k} = 1$ is even) then the multiplication of all such B at the end of dealing with the off-diagonal elements gives a value of 0 at $V'_{k,k}$. Hence we need not do anything to change this value of $V'_{k,k}$. On the other hand if the number of such l is odd then $V'_{k,k} = 1$ and we need to multiply R' by the symplectic matrix for $e^{i\frac{\pi}{4}\tau_{\bar{b}}}$ where b is a length $2n$ vector that is non-zero only at k. As discussed above, this corresponds to a single-qubit rotation about σ_Z.

In the complementary case, if $V_{k,k} = 1$ and the number of such l is odd then the multiplication of all such B at the end of dealing with the off-diagonal elements gives a value of 1 at $V'_{k,k}$. Hence we need not do anything to change this value of $V'_{k,k}$. If the number of such l is even then $V'_{k,k} = 0$ and similarly to the above case we need to multiply R' by the symplectic matrix for $e^{i\frac{\pi}{4}\tau_{\bar{b}}}$.

We have now explicitly described how to decompose the symplectic matrix associated with an arbitrary Clifford element into symplectic matrices for 1 and 2-qubit Clifford operations that correspond to Hadamard's, single qubit rotations about σ_Z, two-qubit

rotations about $\sigma_Z \otimes \sigma_Z$, two-qubit permutation operations and $CNOT$ operations. We now need to deal with the evolution of h vectors in this decomposition. We can find the net h vector for the entire decomposition using Theorem 2 of [41] (which tells us how C, h and d transform under composition of Clifford elements). This theorem states the following:

Suppose C_1, h_1, d_1 and C_2, h_2, d_2 represent two Clifford operations $\mathcal{Q}_1$ and $\mathcal{Q}_2$ respectively. Let

$$\begin{aligned} \overline{C_1} &= \begin{pmatrix} C_1 & 0 \\ d_1^T & 1 \end{pmatrix} \\ \overline{U} &= \begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix} \\ \overline{h_1} &= \begin{pmatrix} h_1 \\ 0 \end{pmatrix} \end{aligned} \tag{E.31}$$

and similarly for the matrices associated with $\mathcal{Q}_2$. Then the composition $\mathcal{Q}_2 \circ \mathcal{Q}_1$ has associated matrices

$$\begin{aligned} \overline{C_{21}} &= \overline{C_2}\ \overline{C_1} \\ \overline{h_{21}} &= \overline{h_1} + \left(\overline{C_1}\right)^T \overline{h_2} + \mathrm{diag}\left(\left(\overline{C_1}\right)^T \mathrm{Lower}\left(\left(\overline{C_2}\right)^T \overline{U}\ \overline{C_2}\right) \overline{C_1}\right). \end{aligned} \tag{E.32}$$

Now the h matrices associated with all of the one and two-qubit Clifford operations listed above are equal to 0. For the Hadamard, two-qubit permutations and CNOT operations this follows immediately from Examples 2,4,5 which explicitly state that $h = 0$ in these cases. For the one and two-qubit $\frac{\pi}{2}$ rotations this follows from Example 7 which states that $h = C^T U a$. However since a is non-zero only in the first n of its $2n$ entries, $Ua = 0$ and so $h = 0$. Hence once we have decomposed our original symplectic matrix into a product of one and two qubit symplectic matrices we can calculate how the h vectors are transformed according to the simpler composition law:

$$\overline{h_{21}} = \mathrm{diag}\left(\left(\overline{C_1}\right)^T \mathrm{Lower}\left(\left(\overline{C_2}\right)^T \overline{U}\ \overline{C_2}\right) \overline{C_1}\right). \tag{E.33}$$

In total then, for our original Clifford element $\mathcal{Q}$ with symplectic matrix $C \in \mathbb{Z}_2^{2n \times 2n}$ and vector $h \in \mathbb{Z}_2^{2n \times 1}$ we can decompose C into a product of symplectic matrices C_k,...,C_1

in $\mathbb{Z}_2^{2n\times 2n}$ that represent the one and two-qubit Clifford operations mentioned above. Each C_i has an associated vector $h_i \in \mathbb{Z}_2^{2n\times 1}$ and we denote the Clifford element associated with C_i and h_i by $\mathcal{Q}_i$. While $C = \prod_{i=1}^{k} C_i$, the net vector obtained from composing the h_i according to Eq. (E.33), call it h', will in general not be equal to h. Hence $\mathcal{Q} \neq \circ_{i=1}^{k} \mathcal{Q}_i$. However the Clifford operation $\mathcal{Q}' := \circ_{i=1}^{k} \mathcal{Q}_i$ (which as noted is associated to $\prod_{i=1}^{k} C_i$ and h'), maps the $2n$ standard generators of $\mathcal{P}_n$ in the same manner as $\mathcal{Q}$ up to negative signs.

However since we have both h and h' we note that $\tilde{h} = h + h'$ is non-zero in an entry if and only if $\mathcal{Q}$ and $\mathcal{Q}'$ disagree on a negative sign. Suppose now that $\tilde{h}_k \neq 0$ for $k \leq n$. Then $\mathcal{Q}$ and $\mathcal{Q}'$ disagree on the negative sign of the image of $\mathbb{1}_2^{\otimes(k-1)} \otimes X \otimes \mathbb{1}_2^{\otimes(n-k)}$. Hence acting the single-qubit Pauli operation with unitary $\mathbb{1}_2^{\otimes(k-1)} \otimes Z \otimes \mathbb{1}_2^{\otimes(n-k)}$ before $\prod_{i=1}^{k} C_i$ maps $\mathbb{1}_2^{\otimes(k-1)} \otimes X \otimes \mathbb{1}_2^{\otimes(n-k)}$ to $-\mathbb{1}_2^{\otimes(k-1)} \otimes X \otimes \mathbb{1}_2^{\otimes(n-k)}$ and so fixes up this negative sign. A symmetric argument holds for the case $k > n$ and so we can see that if we act P on $\tilde{h}$ (which switches the top and bottom halves of the length $2n$ vector) we obtain a description of all of the single-qubit σ_X and σ_Z operations that need to be performed before $\prod_{i=1}^{k} C_i$ in order to fix up all of the negative signs.

Thus if we define $\mathcal{Q}_0 = \tau_{P\tilde{h}} = \tau_{P(h+h')}$ we get $C_0 = \mathbb{1}_{2n}$, $h_0 = P^2(h + h') = h + h'$ (ie. see Example 1) and $\mathcal{Q} = \circ_{i=0}^{k} \mathcal{Q}_i$. Thus $C = \prod_{i=0}^{k} C_i$ and h is the net vector obtained from the composition. To see this we can explicitly use Eq. (E.2). Let $C_2 = C$ and C_1 be the symplectic matrix associated with $\mathcal{Q}_0$ which in this case is $\mathbb{1}_{2n}$ (see Example 1). As well, set $h_2 = h'$ and $h_1 = h' + h$. Then, noting that d_1 is the zero vector,

$$\begin{aligned} h_{21} &= h' + h + h' + \text{diag}\left(\text{Lower}\left(\left(\overline{C_2}\right)^T \overline{U}\, \overline{C_2}\right)\right) \\ &= h \end{aligned} \tag{E.34}$$

since the diagonal elements of a strictly lower triangular matrix are 0. Thus we have finally obtained the decomposition of a Clifford element into one and two-qubit operations. We now collect all of these results into the following main result:

Main Result: Let $\mathcal{Q}$ be a Clifford operation with symplectic matrix C and vector h. Then $\mathcal{Q}$ can be realized by a sequence of one and two-qubit Clifford operations that consists of the following six rounds of operations:

1. An initial round of single-qubit Pauli operators,

2. Applying a sequence of $CNOT$ and two-qubit permutation operations,

3. Applying a sequence of $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$ followed by a sequence of $\frac{\pi}{2}$ rotations about σ_Z,

4. Applying Hadamard operations,

5. Applying a sequence of $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$ followed by a sequence of $\frac{\pi}{2}$ rotations about σ_Z,

6. Applying a final round of $CNOT$ and two-qubit permutation operations.

Note that the operations within each of the rounds 3, 4 and 5 all commute and can be performed in any order. The operations in 2 and 6 are ordered according to the sequence of operations required to bring matrices 2 and 4 in Eq. (E.22) to RREF, and hence order must be preserved. Lastly, there is no order of operations to worry about in round 1.

In the benchmarking protocol one has to find the inverse of a Clifford element for the final deterministic gate in each sequence. This can be done using the following result from [41]:

If $\overline{C_1}$ and $\overline{h_1}$ represent a Clifford operation $\mathcal{Q}_1$ then $\mathcal{Q}_2 = \mathcal{Q}_1^{-1}$ has representation:

$$\begin{aligned} \overline{C_2} &= \left(\overline{C_1}\right)^{-1} = \begin{pmatrix} C_1^{-1} & 0 \\ d_1^T C_1^{-1} & 1 \end{pmatrix} = \begin{pmatrix} P C_1^T P & 0 \\ d_1^T P C_1^T P & 1 \end{pmatrix} \\ \overline{h_2} &= \left(\overline{C_1}\right)^{-T} \overline{h_1} + \mathrm{diag}\left(\left(\overline{C_1}\right)^{-T} \mathrm{Lower}\left(\left(\overline{C_2}\right)^T \overline{U} \left(\overline{C_2}\right)\right) \left(\overline{C_1}\right)^{-1}\right). \end{aligned} \tag{E.35}$$

In many cases one would like to have a decomposition of a Clifford element into a minimal generating set for the Clifford group. Such a set is given by $G_n := \{H, S, CNOT\}$ which consists of Hadamard's (H) and phase gates (S) on each qubit, as well as $CNOT$ gates on all pairs of qubits. It is straightforward to see there are $n^2 + n$ elements in this set. In order to obtain a sequence of elements chosen only from G_n we need to find decompositions of single qubit Pauli's, $e^{\frac{i\pi}{4}Z}$ and $e^{\frac{i\pi}{4}Z\otimes Z}$ in terms of G_n. First, we have $Z = S^2$ and so since $HZH = X$, $X = HS^2H$. Hence, Y can be realized by $ZX = S^2HS^2H$ (global phases can be ignored here). Next, note that in the standard basis,

$$e^{\frac{i\pi}{4}Z} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix} \tag{E.36}$$

and

$$SHS = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \tag{E.37}$$

Hence,

$$e^{\frac{i\pi}{4}Z} = HSHSH = \frac{1}{\sqrt{2}} \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}. \tag{E.38}$$

Lastly, notice that

$$e^{\frac{i\pi}{4}Z\otimes Z} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1+i & 0 & 0 & 0 \\ 0 & 1-i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 0 & 1+i \end{pmatrix} \tag{E.39}$$

and

$$\mathcal{I} \otimes e^{\frac{i\pi}{4}Z} = \mathcal{I} \otimes HSHSH = \frac{1}{\sqrt{2}} \begin{pmatrix} 1+i & 0 & 0 & 0 \\ 0 & 1-i & 0 & 0 \\ 0 & 0 & 1+i & 0 \\ 0 & 0 & 0 & 1-i \end{pmatrix}. \tag{E.40}$$

Hence we need only switch the bottom two diagonals. This can be done by conjugating the Kraus operator for $CNOT_1$ where the subscript "1" indicates the first qubit is the control qubit. This gives,

$$e^{\frac{i\pi}{4}Z\otimes Z} = CNOT_1 \left(\mathcal{I} \otimes HSHSH\right) CNOT_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1+i & 0 & 0 & 0 \\ 0 & 1-i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 0 & 1+i \end{pmatrix} \tag{E.41}$$

Appendix F

Randomized Benchmarking: Experimental Protocol

In this section we provide a detailed step-by-step explanation of how to perform the randomized benchmarking protocol outlined in Chapter 2. We also provide various numerical routines that can be used in the different steps of the protocol. The numerical routines rely on the theory from the previous sections regarding the symplectic representation of the Clifford group and how to decompose Clifford elements into a sequence of generators.

F.1 Experimental Protocol For Implementing Randomized Benchmarking

In this section we discuss the experimental protocol for implementing the randomized benchmarking algorithm presented in Chapter. 2. The protocol is as follows:

Step 1: Choose an input stabilizer state $|\psi\rangle$ that is simple to prepare.

Step 2: Choose a set of R positive integers $m_1 <...< m_R$, and define $M = m_R + 1$.

-Each $m_j + 1$ will correspond to a different sequence length for which the benchmarking protocol is performed (the sequence length corresponds to the horizontal axis of the fidelity decay curve).

-M is the maximum sequence length.

Step 3: For each $j \in \{1, ..., R\}$ repeat the following sub-steps K times, where K is defined in Eq. (2.110) (K depends on the desired confidence interval/accuracy of the data points):

Sub-step a) Choose m_j gates $\mathcal{C}_{i_1}$,...,$\mathcal{C}_{i_{m_j}}$ uniformly at random from the Clifford group (Matlab code for producing a uniformly random Clifford element is available upon request [1])

Sub-step b) Determine a final inverse gate $\mathcal{C}_{i_{m_j+1}}$ that maps the state

$$\mathcal{C}_{i_{m_j}} \circ ... \circ \mathcal{C}_{i_1}(|\psi\rangle\langle\psi|) \tag{F.1}$$

back to $|\psi\rangle\langle\psi|$. For small numbers of qubits, $\mathcal{C}_{i_{m_j+1}}$ can be found by just calculating $\mathcal{C}_{i_{m_j+1}} = \mathcal{C}_{i_1}^{\dagger} \circ ... \circ \mathcal{C}_{i_{m_j}}^{\dagger}$ (Matlab code for finding the inverse of a Clifford element is available upon request).

Sub-sub-step i) For each i_k, where $k \in \{1, ..., m_j + 1\}$, decompose $\mathcal{C}_{i_k}$ into the sequence of five Clifford operations $T_1^{i_k}$,...,$T_5^{i_k}$ as outlined in the previous section and Ref. [41] (Matlab code for this decomposition is available upon request).

Sub-sub-step ii) Decompose each of $T_1^{i_k}$,...,$T_5^{i_k}$ into the sequence of generators from the previous section and Ref. [41] (Matlab code for these decompositions is available upon request)

Sub-sub-step iii) For each i_k, find the necessary sequence of Pauli operations that produce the correct h vector (Matlab code that produces this sequence is available upon request).

Sub-step c): Implement all of the operations obtained starting from Sub-step a) in the correct order on $|\psi\rangle\langle\psi|$ and measure the probability of obtaining the measurement result "ψ" from a measurement whose basis contains $|\psi\rangle$.

[1] email: emagesan@gmail.com

Step 4: Average over the K values one obtains from repeating Step 3 to obtain $F_{\text{seq}}(m_j, \psi)$.

Step 5: Plot $F_{\text{seq}}(m_j, \psi)$ as a function of m_j to obtain a fidelity decay curve.

Step 6: Fit the fidelity decay curve to either the zeroth order model given by Eq. (2.34) or the first order model of Eq. (2.54) to obtain an estimate of p.

Appendix G

Partial list of abbreviations

Table G.1: Partial list of abbreviations used.

Abbreviation	Full Name
a-sym(k, d)	anti-symmetric subspace of k subsystems each of dimension d
Clif_n	Clifford group on n qubits
$C(\Lambda)$	Choi matrix of a quantum operation Λ
CNOT	Controlled-NOT gate
$\mathbb{CP}^{d-1}$	Complex projective space of a d-dimensional space
$\mathcal{F}_{\mathcal{E},\mathcal{U}}$	Gate fidelity between a quantum operation $\mathcal{E}$ and unitary $\mathcal{U}$
H	Hadamard gate
$J(\Lambda)$	Jamiolkowski state of a quantum operation Λ
Λ_d	Unique depolarizing channel having the same average fidelity as Λ
n	number of qubits
$\mathcal{P}_n$	Pauli group on n qubits
POVM	Positive operator-valued measure
QEC	Quantum error correction
QPT	Quantum process tomography
RB	Randomized benchmarking
S	Phase gate
sym(k, d)	symmetric subspace of k subsystems each of dimension d
$U(d)$	Unitary group on a d-dimensional Hilbert space

References

[1] S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, 2004. 40

[2] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, 1997. 3, 36

[3] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the 30'th annual ACM symposium on theory of computing*, Dallas, TX, 1998. ACM. 8, 50

[4] D. Aharonov, A. Kitaev, and J. Preskill. Fault-tolerant quantum computation with long-range correlated noise. *Phys. Rev. Lett.*, 96(5):050504, Feb 2006. 135

[5] P. Aliferis, F. Brito, D.P. DiVincenzo, J. Preskill, M. Steffen, and B.M. Terhal. Fault-tolerant computing with biased-noise superconducting qubits: a case study. *New J. Phys.*, 11(1):013061, 2009. 136

[6] P. Aliferis and A.W. Cross. Subsystem fault tolerance with the bacon-shor code. *Phys. Rev. Lett.*, 98(22):220502, May 2007. 135

[7] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White. Ancilla-assisted quantum process tomography. *Phys. Rev. Lett.*, 90:193601, May 2003. 5, 11

[8] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *Proceedings of Complexity '07*, pages 129–140, 2007. 83

[9] T. R. Beals, J. Vala, and K. B. Whaley. Scalability of quantum computation with addressable optical lattices. *Phys. Rev. A*, 77(5):052309, May 2008. 136

[10] S. Beigi and R. Koenig. Simplified instantaneous non-local quantum computation with applications to position-based cryptography. *New J. Phys.*, 13(9):093036, 2011. 37, 119

[11] A. Bendersky, F. Pastawski, and J.P. Paz. Selective and efficient estimation of parameters for quantum process tomography. *Phys. Rev. Lett.*, 100:190403, 2008. 6, 51

[12] I. Bengtsson and K. Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, Cambridge, UK, 2006. 80, 118, 123

[13] C.H. Bennett and G. Brassard. Quantum cryptography: Public-key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, Bangalore, India, 1984. 1

[14] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, and W.K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824, 1996. 91, 133, 149

[15] M.J. Biercuk, H. Uys, A.P. VanDevender, N. Shiga, W.M. Itano, and J.J. Bollinger. High-fidelity quantum control using ion crystals in a penning trap. *Quantum Inf. Comput.*, 9(11):0920, 2009. 7, 49

[16] F. Bloch. Nuclear induction. *Phys. Rev.*, 70:460–474, 1946. 58, 123

[17] R. Blume-Kohout, H.K. Ng, D. Poulin, and L. Viola. The structure of preserved information in quantum processes. *Phys. Rev. Lett*, 100:030501, 2008. 51

[18] V. Bouchiat, D. Vion, P. Joyez, D. Esteve, and M.H. Devoret. Quantum coherence with a single cooper pair. *Phys. Scr. A*, T76:165–170, 1998. 2

[19] P. S. Bourdon and H. T. Williams. Unital quantum operations on the bloch ball and bloch region. *Phys. Rev. A*, 69(2):022314, Feb 2004. 123

[20] P.O. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan. On universal and fault-tolerant quantum computing: A novel basis and a new constructive proof of universality for shor's basis. In *Proceedings of the 40'th Annual Symposium on Foundations of Computer Science (FOCS)*, 1999. 126, 127

[21] S. Bravyi and A. Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Phys. Rev. A*, 71:022316, Feb 2005. 46, 127, 133

[22] G.K. Brennen, C.M. Caves, P.S. Jessen, and I.H. Deutsch. Quantum logic gates in optical lattices. *Phys. Rev. Lett.*, 82:1060–1063, Feb 1999. 2

[23] H.P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, 2002. 85

[24] K. R. Brown, A. C. Wilson, Y. Colombe, C. Ospelkaus, A. M. Meier, E. Knill, D. Leibfried, and D. J. Wineland. Single-qubit-gate error below 10^{-4} in a trapped ion. *Phys. Rev. A*, 84:030303, Sep 2011. ix, 43, 45, 49

[25] A.M. Buckner, J.B. Bruckner, and B.S. Thomson. *Real Analysis*. Prentice-Hall, 1997. 154

[26] C.K. Burrell. Geometry of generalized depolarizing channels. *Phys. Rev. A*, 80:042330, 2009. 104

[27] A.R. Calderbank and P.W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1098, 1996. 3, 46, 133

[28] D.E. Chang, L.M.K. Vandersypen, and M. Steffen. Nmr implementation of a building block for scalable quantum computation. *Chem. Phys. Lett.*, 338(46):337 – 344, 2001. 136

[29] M.D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and Its Applications*, 10:285–290, 1975. 32, 98

[30] M.D. Choi and D.W. Kribs. A method to find quantum noiseless subsystems. *Physical Review Letters*, 96:050501, 2006. 131

[31] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf. Optimized driving of superconducting artificial atoms for improved single-qubit gates. *Phys. Rev. A*, 82:040305, Oct 2010. 7, 49

[32] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf. Randomized benchmarking and process tomography for gate errors in a solid-state qubit. *Phys. Rev. Lett.*, 102:090502, 2009. 7, 49

[33] I.L. Chuang and M.A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.*, 44(11):2455, 1997. 3, 5, 11, 51, 150

[34] I.J. Cirac and P. Zoller. Quantum computations with cold trapped ions. *Phys. Rev. Lett.*, 74:4091, 1995. 2

[35] E. Collini, C.Y. Wong, K.E. Wilk, P.M.G. Curmi, P. Brumer, and G.D. Scholes. Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature*, 463(7281):644–647, 2010. 2

[36] D. G. Cory, M. D. Price, W. Maas, E. Knill, R. Laflamme, W. H. Zurek, T. F. Havel, and S. S. Somaroo. Experimental quantum error correction. *Phys. Rev. Lett.*, 81:2152–2155, Sep 1998. 2

[37] D.G. Cory, A.F. Fahmy, and T.F Havel. Nuclear magnetic resonance spectroscopy: an experimentally accessible paradigm for quantum computing. In *Proceedings of the 4th Workshop on Physics and Computation*, Boston, MA, 1996. 2

[38] M.P. da Silva, O. Landon-Cardinal, and D. Poulin. Practical characterization of quantum devices without tomography. arXiv:1104.3835v3, June 2011. 6, 11

[39] C. Dankert. Efficient simulation of random quantum states and operations. Ph.D. Thesis, arXiv:quant-ph/0512217v2, Dec 2005. 39

[40] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their applications to fidelity estimation. *Phys. Rev. A*, 80:012304, 2009. 8, 51, 91, 104, 126, 146, 147, 148, 149

[41] J. Dehaene and B. De Moor. Clifford group, stabilizer states, and linear and quadratic operations over gf(2). *Phys. Rev. A*, 68(4):042318, Oct 2003. 39, 40, 164, 165, 166, 167, 168, 170, 171, 173, 176, 178, 181

[42] P. Delsarte, J.M. Goethals, and J.J. Seidel. Spherical codes and designs. *Geometriae Dedicata: Constructions and Applications*, 6, 1997. 145

[43] D. Deutsch. Quantum theory, the church-turing principle and the universal quantum computer. *Proc. Roy. Soc. Lond. A*, 400:97–117, 1985. 1

[44] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *Information Theory, IEEE Transactions on*, 51(1):44 –55, Jan. 2005. 144

[45] D.P. DiVincenzo. The physical implementation of quantum computation. *Fortschritte der Physik*, 48(9-11):771–783, 2000. 1, 2

[46] J. Emerson, R. Alicki, and K. Zyczkowski. Scalable noise estimation with random unitary operators. *J. Opt. B: Quantum and Semiclassical Optics*, 7(10):S347–S352, 2005. 6, 7, 8, 51, 60, 91, 150

[47] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D.G. Cory, and R. Laflamme. Symmetrized characterization of noisy quantum processes. *Science*, 317(5846):1893–1896, 2007. 6, 11, 51, 91

[48] G.S. Engel, T.R. Calhoun, E.L. Read, T. Ahn, T. Mancal, Y. Cheng, R.E. Blankenship, and G.R. Fleming. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature*, 446(7137):782–786, 2007. 2

[49] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum computation by adiabatic evolution, 2000. quant-ph/0001106. 1

[50] S.T. Flammia and Y-K. Liu. Direct fidelity estimation from few pauli measurements. *Phys. Rev. Lett.*, 106:230501, Jun 2011. 6, 11

[51] C. A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. PhD thesis, University of New Mexico, Albuquerque, New Mexico, 2006. Preprint quant-ph 9601020. 50, 106

[52] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inf. Th.*, 45(4):1216–1227, 1999. 114

[53] A. Fujiwara and P. Algoet. Affine parameterization of quantum channels. In *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, page 87, August 1998. 123

[54] H. Georgi. *Lie Algebras in Particle Physics*. Westview Press, U.S.A, second edition, 1999. 104, 124

[55] N.A. Gershenfeld and I.L. Chuang. Bulk spin-resonance quantum computation. *Science*, 275(5298):350–356, 1997. 2

[56] A. Gilchrist, N.K. Langford, and M.A. Nielsen. Distance measures to compare real and ideal quantum processes. *Phys. Rev. A*, 71:062310, 2005. 50, 51, 89, 109, 120

[57] D. Gottesman. Stabilizer codes and quantum error correction. Ph.D. Thesis, arXiv:quant-ph/9705052, May 1997. 6, 126, 127, 128, 132, 164

[58] D. Gottesman and I.L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402:390–393, 1999. 1

[59] M. Gromov and V.D. Milman. A topological application of the isoperimetric inequality. *American Journal of Mathematics*, 105(4):843, 1983. 156

[60] L.K. Grover. A fast quantum mechanical algorithm for database search. In *28th Annual ACM Symposium on the Theory of Computing*, 1996. 1

[61] R. Hanson, O. Gywat, and D. D. Awschalom. Room-temperature manipulation and decoherence of a single spin in diamond. *Phys. Rev. B*, 74:161203, Oct 2006. 2

[62] A.W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for solving linear systems of equations. *Phys. Rev. Lett.*, 103:150502, 2009. 1

[63] A.W. Harrow and R.A. Low. Efficient quantum tensor product expanders and k-designs. In *Proceedings of the 12th International Workshop and 13th International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, APPROX '09 / RANDOM '09, pages 548–561, Berlin, Heidelberg, 2009. Springer-Verlag. 92

[64] M.B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255, 2009. 92, 143

[65] P. Hayden, D. Leung, and A. Winter. Aspects of generic entanglement. *Comm. Math. Phys.*, 265(1):95, 2006. 92

[66] H.F. Hofmann. Complementary classical fidelities as an efficient criterion for the evaluation of experimentally realized quantum operations. *Phys. Rev. Lett.*, 94(16):160504, Apr 2005. 91

[67] J.A. Holbrook, D.W. Kribs, and R. Laflamme. Noiseless subsystems and the structure of the commutant in quantum error correction. *Quant. Inf. Proc.*, 2:381–419, 2004. 131

[68] A. Holevo. *Statistical problems in quantum physics*, volume 330 of *Lecture Notes in Mathematics*. Springer Berlin / Heidelberg, 1973. 10.1007/BFb0061483. 137

[69] A.S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Th.*, 44(1):269 –273, January 1998. 142

[70] A.S. Holevo and V. Giovannetti. Quantum channels and their entropic characteristics. *Rep. Prog. Phys.*, 75(4):046001, 2012. 136

[71] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1990. 82

[72] R. Jain, Z. Ji, S. Upadhyay, and J. Watrous. Qip = pspace. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 573–582, New York, NY, USA, 2010. ACM. 1

[73] N. Johnston and D.W Kribs. Quantum gate fidelity in terms of choi matrices. *J. Phys. A: Mathematical and Theoretical*, 44(49):495303, 2011. 91, 93

[74] N. Johnston and D.W Kribs. A family of norms with applications in quantum information theory. *J. Math. Phys.*, 51(082202), 2019. 93

[75] R. Jozsa. Quantum algorithms and the fourier transform. arXiv:quant-ph/9707033, 1997. 127

[76] R. Jozsa. Quantum factoring, discrete logarithms, and the hidden subgroup problem. *Computing in Science and Engineering*, 03:34–43, 2001. 127

[77] A.Y. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303:2–30, 1997. 1

[78] A.Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52:1191–1249, 1997. 36, 50, 111

[79] A.Y. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, Boston, MA, USA, 2002. 8

[80] E. Knill. Quantum computing with realistically noisy devices. *Nature*, 434:39–44, 2005. 135

[81] E. Knill. Protected realizations of quantum information. *Phys. Rev. A*, 74:042301, 2006. 131

[82] E. Knill, R. Laflamme, and L. Viola. Theory of quantum error correction for general noise. *Phys. Rev. Lett.*, 84:2525–2528, Mar 2000. 3

[83] E. Knill, R. Laflamme, and W. Zurek. Resilient quantum computation: error models and thresholds. *Proc. R. Soc. Lond. A*, 454:365–384, 1997. 3

[84] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, 2008. 7, 8, 48, 49, 91, 92

[85] K. Kraus. *States, Effects and Operations*. Springer-Verlag, Berlin, 1983. 98

[86] D.W. Kribs, R. Laflamme, and D. Poulin. Unified and generalized approach to quantum error correction. *Phys. Rev. Lett.*, 94:180501, May 2005. 128, 130

[87] D.W. Kribs, A. Pasieka, M. Laforest, C. Ryan, and M. Silva. Research problems on numerical ranges in quantum computing. *Linear and Multilinear Algebra*, 57:491–502, 2009. 51, 93

[88] D.W. Kribs and R.W. Spekkens. Quantum error correcting subsystems are unitarily recoverable subsystems. *Phys. Rev. A*, 74:042329, 2006. 131

[89] R. Laflamme, C. Miquel, J.P. Paz, and W.H. Zurek. Perfect quantum error correcting code. *Phys. Rev. Lett.*, 77(1):198–201, Jul 1996. 133

[90] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001. 86, 156, 158, 163

[91] J.M. Lee. *Introduction to Smooth Manifolds*. Springer, 2002. 154

[92] B. Levi, C. C. Lopez, J. Emerson, and D. G. Cory. Efficient error characterization in quantum information processing. *Phys. Rev. A*, 75:022314, 2007. 8

[93] J.E. Levy et al. The impact of classical electronics on a solid-state logical qubit memory. In *Proceedings of the twenty-first annual symposium on Parallelism in algorithms and architectures*, Calgary, Canada, 2009. 136

[94] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996. 1

[95] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55(3):1613–1622, Mar 1997. 144

[96] D. Loss and D.P. Divincenzo. Quantum computation with quantum dots. *Phys. Rev. A*, 57:120–126, 1998. 2

[97] E. Magesan. Depolarizing behavior of quantum channels in higher dimensions. *Quant. Inf. Comp.*, 11(5):0466–0484, 2011. 4, 48, 51, 118

[98] E. Magesan, R. Blume-Kohout, and J. Emerson. Gate fidelity fluctuations and quantum process invariants. *Phys. Rev. A*, 84(1):012309, Jul 2011. 4, 48, 51, 118

[99] E. Magesan, J. M. Gambetta, and J. Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106(18):180504, May 2011. 4, 8, 47, 91, 150

[100] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, Apr 2012. 4, 8

[101] V. D. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Springer-Verlag, 1980. Lecture Notes in Mathematics-1200. 88, 156, 162, 163

[102] M. Mohseni and D. A. Lidar. Direct characterization of quantum dynamics. *Phys. Rev. Lett.*, 97:170501, Oct 2006. 5

[103] T. Monz, P. Schindler, J.T. Barreiro, M. Chwalla, D. Nigg, W.A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt. 14-qubit entanglement: Creation and coherence. *Phys. Rev. Lett.*, 106:130506, Mar 2011. 47

[104] O. Moussa, M. Silva, C.A. Ryan, and R. Laflamme. Practical experimental certification of computational quantum gates via twirling. arXiv:1112.4505v1, Dec 2011. 6, 11

[105] J.R. Munkres. *Topology*. Prentice Hall Inc., second edition edition, 2000. 153

[106] H.K Ng and J. Preskill. Fault-tolerant quantum computation versus gaussian noise. *Phys. Rev. A*, 79(3):032318, Mar 2009. 135

[107] M.A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Phys. Lett. A*, 303:249–252, 2002. 51, 59, 60, 118, 150

[108] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Information*. Cambridge University Press, Cambridge, UK, 2000. 82, 97, 100, 107, 108, 118, 119, 127, 129, 132, 133

[109] D.K.L. Oi, S.J. Devitt, and L.C.L. Hollenberg. Scalable error correction in distributed ion trap computers. *Phys. Rev. A*, 74(5):052313, Nov 2006. 136

[110] S. Olmschenk, R. Chicireanu, K. D. Nelson, and J. V. Porto. Randomized benchmarking of atomic qubits in an optical lattice. *New J. Phys.*, 12:113007, 2010. 7, 49

[111] R. Orus and R. Tarrach. Weakly entangled states are dense and robust. *Phys. Rev. A*, 70:050101, 2004. 96

[112] V. Paulsen. *Completely Bounded Maps and Operator Algebras*, volume 78. Cambridge University Press, UK, 2002. 98, 110

[113] L.H. Pedersen, N.M. Møller, and K.Mølmer. The distribution of quantum fidelities. *Phys. Lett. A*, 372(47):7028–7032, 2008. 51

[114] V. Pestov. On the geometry of similarity search: Dimensionality curse and concentration of measure. *Inf. Proc. Lett.*, 73:47–51, 2000. 156, 163

[115] M.B. Plenio and S.F. Huelga. Quantum dynamics of bio-molecular systems in noisy environments. *Proc. Chem.*, 3:248, 2011. 3

[116] S. Popescu, A.J. Short, and A. Winter. Entanglement and the foundations of statistical mechanics. *Nature Physics*, 2:754–758, 2006. 92

[117] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Phys. Rev. Lett.*, 78:390–393, 1997. 3, 5, 51

[118] J. Preskill. Fault tolerant quantum computation. arXiv:quant-ph/9712048, 1997. 3

[119] Z. Puchala, J.A. Miszczak, P. Gawron, and B. Gardas. Experimentally feasible measures of distance between quantum operations. *Quant. Inf. Proc. (in press)*, 2010. arXiv:0911.0567. 51

[120] R. Raussendorf and H.J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, May 2001. 1

[121] R. Raussendorf and J. Harrington. Fault-tolerant quantum computation with high threshold in two dimensions. *Phys. Rev. Lett.*, 98(19):190504, May 2007. 135

[122] J.M. Renes, R. Blume-Kohout, A. J. Scott, and C.M. Caves. Symmetric informationally complete quantum measurements. *J. Math. Phys.*, 45:2171, 2004. 83, 151

[123] M.B. Ruskai, Stanislaw Szarek, and Elisabeth Werner. An analysis of completely-positive trace-preserving maps. *Linear Algebra and its Applications*, 347:159–187, 2002. 58, 123

[124] C.A. Ryan, M. Laforest, and R. Laflamme. Randomized benchmarking of single and multi-qubit control in liquid-state nmr quantum information processing. *New J. Phys.*, 11:013034, 2009. 7, 49

[125] M.F. Sacchi. Minimum error discrimination of pauli channels. *J. Opt. B*, 7(S333), 2005. 31

[126] C.T. Schmiegelow, A. Bendersky, M.A. Larotonda, and J.P. Paz. Selective and efficient quantum process tomography without ancilla. *Phys. Rev. Lett.*, 107:100502, Sep 2011. 6, 11

[127] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56(1):131–138, Jul 1997. 142

[128] P.D. Seymour and T. Zaslavsky. Averaging sets: A generalization of mean values and spherical designs. *Advances in Mathematics*, 52:213–240, 1984. 146

[129] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White. Efficient measurement of quantum dynamics via compressive sensing. *Phys. Rev. Lett.*, 106:100401, Mar 2011. 5

[130] P.W. Shor. The quantum channel capacity and coherent information. Lecture notes, MSRI Workshop on Quantum Computation, 2002. Available online at http://www.msri.org/publications/ln/msri/2002/. 144

[131] P.W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35'th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 124–134, Los Alamitos, CA, 1994. IEEE Press. 1, 127

[132] P.W. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:R2493, 1995. 3

[133] P.W. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246:453–472, 2004. 10.1007/s00220-003-0981-7. 143

[134] M. Silva, E. Magesan, D.W. Kribs, and J. Emerson. Scalable protocol for identification of correctable codes. *Phys. Rev. A*, 78:012347, 2008. 6, 51, 91

[135] A. Steane. Multiple particle interference and quantum error correction. *Proc. Roy. Soc. of London A*, 452(1954):2551–2577, 1996. 3, 46, 133

[136] A.M. Steane. Overhead and noise threshold of fault-tolerant quantum error correction. *Phys. Rev. A*, 68(4):042322, Oct 2003. 135

[137] W.F. Stinespring. Positive functions on C*-algebras. *Proc. Amer. Math. Soc.*, pages 211–216, 1955. 98

[138] K.M. Svore, B.M. Terhal, and D.P. DiVincenzo. Local fault-tolerant quantum computation. *Phys. Rev. A*, 72(2):022317, Aug 2005. 135

[139] J. M. Taylor et al. Fault-tolerant architecture for quantum computation using electrically controlled semiconductor spins. *Nature*, 1:177–183, 2005. 136

[140] B.M. Terhal and G. Burkard. Fault-tolerant quantum computation for local non-markovian noise. *Phys. Rev. A*, 71(1):012336, Jan 2005. 135

[141] L. Viola, E. Knill, and S. Lloyd. Dynamical decoupling of open quantum systems. *Phys. Rev. Lett.*, 82:2417–2421, Mar 1999. 128

[142] J. Watrous. Notes on super-operator norms induced by schatten norms. *Quant. Inf. Comput.*, 5(1):058–067, 2005. 28, 81, 111, 117

[143] J. Watrous. Quantum computational complexity, 2009. Encyclopedia of Complexity and System Science. 1

[144] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11), 2009. 31, 32, 46

[145] X. Zhou, D.W. Leung, and I.L. Chuang. Methodology for quantum logic gate constructions. *Phys. Rev. A*, 62:052316, 2000. 127