A Polar Decomposition for Quantum Channels: Theory and Applications

by

Matthew Alexander

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

Waterloo, Ontario, Canada, 2019

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

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

Statement of Contributions

1. Introduction

    - The arguments of section 1.1 are by Matthew Alexander. All other information in chapter 1 is common knowledge.

2. Quantum Channels

    - All information in chapter 2 is common knowledge or the original source has been cited.

3. The Leading Kraus Approximation

    - The results from the beginning of chapter 3 to the end of section 3.4 were the product of a collaboration between Arnaud Carignan-Dugas and Matthew Alexander.
    - The definitions and examples of bitter channels in section 3.5 were provided by Arnaud Carignan-Dugas. The discussions of bitter channels and their intuition were provided by Matthew Alexander. Figure 3.5 was provided by Matthew Alexander. Subsection 3.5.3 on equability was the result of a collaboration between Arnaud Carignan-Dugas and Matthew Alexander.
    - The results of section 3.6 were the product of a collaboration between Arnaud Carignan-Dugas and Matthew Alexander.
    - The results of sections 3.7 to 3.9 were originally proved by Arnaud Carignan-Dugas. An alternative proof of the closure of decoherence limited channels in section 3.9 was found by Matthew Alexander and is presented here.

4. Conclusions and Future Directions

    - The future directions presented are a result of discussions between Arnaud Carignan-Dugas and Matthew Alexander.

Abstract

Every potential benefit of quantum computers would be lost without methods to protect the computations from error. Quantum channels provide a framework for understanding error in quantum systems. In general, for a system of size d, a quantum channel may require as many as $\sim d^4$ parameters to characterize it. We introduce the leading Kraus approximation, a simplification which reduces the number of parameters to $\sim d^2$, while accurately approximating two figures of merit important to the experimentalist: the unitarity and average process fidelity. Additionally, applying the leading Kraus approximation declutters investigations into the set of quantum channels. When applied, a natural decomposition of channels arises, separating behaviour into coherent and decoherent contributions. We find that eliminating the coherent contribution provides the greatest increase to the fidelity, while decoherent processes provide a generalization of depolarizing channels.

Acknowledgements

I want to thank my supervisor, Joseph Emerson, for his guidance and for providing me with an opportunity to explore the field of quantum information to my heart's content. I would also like to thank all of the members of my committee — Joel Wallman, Ray Laflamme, and Crystal Senko — for all of the time that they have spent in connection with this thesis.

I want to thank all of the members of my research group, especially Arnaud Carignan-Dugas, for his continued insight, advice, and friendship.

Finally, to all of my friends and family, whose love, support, and faith steadied my head and my progress, you have all of my gratitude and love.

Dedication

To Mother

To Pa

To Lauren

To Arthur

Tice

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Given a pair of six-sided dice, there is an obvious probabilistic algorithm available for approaching any problem whose solution lies between 2 and 12. Unfortunately, this protocol is not much use to us if, in computing $3+4$, we roll our dice and receive 10. Though there is an extended capacity to our dice — we have access to 10 more solutions than if each die stored only one number — we also extend our space of erroneous solutions. Such a probabilistic computing scheme necessitates research into devising algorithms which will produce correct outcomes with as high a likelihood as possible.

This trade-off between capacity and uncertainty characterizes the difference between probabilistic and deterministic classical computing schemes. Quantum mechanics, inherently non-deterministic, gives rise to a form of computing in which probabilities serve in key roles. Quantum computing can, at first pass, be viewed as similar to classical probabilistic computing. However, quantum computations seek to manipulate a subtler resource than classical probabilistic computing has access to: the amplitudes of quantum states; those complex parameters whose lengths correspond to the probabilities of classical computing, and whose directions are hidden from classical control. Quantum algorithms labour to tune these amplitudes appropriately [30, 25] while quantum error correction seeks to repair them [2, 10].

In the background, ever present, noise threatens to try its hand at amplitude manipulation; decohering states and ruining computations. We can investigate noise processes through the mathematics of **quantum channels**. These maps may rely on $\sim d^4$ paramaters, where d is the dimension of our system (which, in turn, is typically of the form

$d = 2^N$, where N is the number of qubits). Quickly then, analyses of these channels require of us massive resources. This work investigates a method of simplifying the study of quantum channels, reducing the number of required parameters to $\sim d^2$. Our proposed **leading Kraus approximation** will be shown (under modest assumptions) to closely approximate several figures of merit of quantum channels which are of importance to the experimentalist.

The remainder of this chapter is dedicated to reviewing some preliminaries from the field of quantum computing. Chapter 2 presents the mathematical formulation of quantum channels, several of their representations, and two quantities which can be used to investigate the channels, viz. the fidelity and unitarity. In chapter 3 we develop the leading Kraus approximation and explore its capabilities and consequences. We conclude by outlining future directions that the leading Kraus approximation can take us in, and in particular, consider how the leading Kraus approximation can be utilized in combination with known protocols to extend their capabilities.

1.2 Quantum States and Operators

The main player on the stage of quantum mechanics is the quantum state — that object which captures the condition of a given quantum system. The worlds of these quantum characters are known as **Hilbert Spaces** — vector spaces endowed with inner products, and closed under the limit of Cauchy sequences of their elements. We will denote Hilbert spaces by $\mathcal{H}$ and the states within by $|\psi\rangle \in \mathcal{H}$. The vector space structure of Hilbert spaces gives us closure under linear combinations of states:

$$\alpha |\psi\rangle + \beta |\phi\rangle \in \mathcal{H}\,,\ \forall \alpha, \beta \in \mathbb{C}\,,\ \forall\, |\psi\rangle\,, |\phi\rangle \in \mathcal{H}. \tag{1.1}$$

The inner product on our space will be denoted with bra-ket notation: $\langle\psi|\phi\rangle \equiv \langle |\psi\rangle\,, |\phi\rangle\rangle$.

The vector space structure also allows us to form bases, $\{|x_i\rangle\}$, in which we can expand any given state: $|y\rangle = \sum_i \alpha_i |x_i\rangle$. We can always choose an orthonormal basis, in which

$$\langle x_i | x_j \rangle = \delta_{ij} = \begin{cases} 1, \text{ if } i = j \\ 0, \text{ if } i \neq j \end{cases}\,. \tag{1.2}$$

The coefficients in expansions of quantum mechanical states, $\sum_i \alpha_i |\psi_i\rangle$, are known as **amplitudes**. We impose a particular restriction on our quantum states: $\langle\psi|\psi\rangle = \sum_i |\alpha_i|^2 = 1$.

Physically, we interpret the $|\alpha_i|^2$'s as probabilities that we associate with the corresponding states, $|x_i\rangle$. The normalization condition on our states then simply assures that our probabilities sum to 1.

We also will concern ourselves with manipulations of our quantum states. Such manipulations will be linear maps from $\mathcal{H} \to \mathcal{H}$, called **quantum operators**. In particular, those manipulations which are physical should send normalized states to normalized states. We require of any quantum operator, $A : \mathcal{H} \to \mathcal{H}$,

$$\langle\psi|A^\dagger A|\psi\rangle = \langle\psi|\psi\rangle \quad , \forall\, |\psi\rangle \in \mathcal{H}. \tag{1.3}$$

We will denote an n-dimensional Hilbert space by $\mathcal{H}^n$. One of the most useful results for the study of Hilbert spaces, is that any finite dimensional Hilbert space, $\mathcal{H}^n$, may be represented by an n-dimensional complex vector space, V^n, with the states in $\mathcal{H}^n$ corresponding to vectors in V^n. Given a finite-dimensional Hilbert space, $\mathcal{H}^n$, we pick an ordered orthonormal basis, $\{|x_i\rangle\}_{i=1}^n$. We then map each element of the basis to a corresponding **elementary vector** in V^n: $|x_i\rangle \to e_i$, where the elementary vectors are given by:

$$e_1 \equiv \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 \equiv \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n \equiv \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{1.4}$$

We can then map any state in the Hilbert space to a corresponding vector in an n-dimensional complex vector space, by extending the mapping linearly:

$$|\psi\rangle = \sum_i \alpha_i |x_i\rangle \to \sum_i \alpha_i e_i \tag{1.5}$$

If we represent $|\psi\rangle$ by a vector, $\psi \in \mathbb{C}^n$, then the dual vector, $\langle\psi|$, is represented by $\psi^\dagger$, where $\dagger$ denotes the conjugate transpose. This is consistent with the notion of the inner product:

$$\langle\psi|\phi\rangle \to \psi^\dagger\phi. \tag{1.6}$$

Similarly, we can represent linear operators $T : \mathcal{H}^n \to \mathcal{H}^m$ by $m \times n$ matrices, $M_T : \mathbb{C}^n \to \mathbb{C}^m$, whose components are determined by

$$\langle x_i| T |x_j\rangle \to e_i^T (M_T) e_j. \tag{1.7}$$

We will only be considering finite dimensional spaces in what follows. Because of this, we will transfer our attention from considering the more abstract ideas of states and operators, to the concrete forms of vectors and matrices. For ease of notation, we shall denote vectors in our complex vector space with the bra-ket notation of the states. In particular, the elementary vectors will be denoted as $e_i \equiv |i\rangle$. The set $\{|i\rangle\}_{i=1}^n$, for an n-dimensional complex vector space, is called the **canonical basis**. We will often write the outer-product of canonical basis elements as **elementary matrices**: $E_{ij} \equiv |i\rangle \langle j|$. We will denote the space of $d \times d$ complex matrices as $\mathbb{M}_d(\mathbb{C})$.

Note: We can use the canonical basis to acquire the columns and elements of a matrix, A:

- $A|i\rangle$ is the ith column of the matrix, A, written, $|a_i\rangle$.
- $\langle i|A|j\rangle$ is the i-jth element of A (ith row, jth column), written, a_{ij}. If we wish to be more explicit, we will write $(A)_{ij}$, instead.

Due to the importance that vectors and matrices play in what follows, we shall review some important results of linear algebra in the remainder of this chapter.

1.3 Matrices

Let $\{|x_i\rangle\}$ be any orthonormal basis. We can express the identity matrix through these vectors as:

$$I = \sum_i |x_i\rangle \langle x_i| . \tag{1.8}$$

We can show that the above relation holds, as follows: Since $\{|x_i\rangle\}$ is an orthonormal basis, we can write any vector in terms of it: $|y\rangle = \sum_i y_i |x_i\rangle$. And so

$$(\sum_i |x_i\rangle \langle x_i|) |y\rangle = \sum_{i,j} y_j |x_i\rangle \langle x_i|x_j\rangle = \sum_j y_j |x_j\rangle = |y\rangle \ \forall |y\rangle . \tag{1.9}$$

It is thus the identity transformation. In fact the reverse holds as well:

Theorem 1.3.1
Let $\{|x_i\rangle\}$ be a set of normalized vectors. Then

$$I = \sum_i |x_i\rangle \langle x_i| \tag{1.10}$$

iff $\{|x_i\rangle\}$ is an orthonormal basis.

Proof. We have already shown one direction. On the other hand, if we have a set of normalized vectors, $\{|x_i\rangle\}$, such that $I = \sum_i |x_i\rangle \langle x_i|$, then

$$1 = \langle x_j|x_j\rangle = \langle x_j|I|x_j\rangle = \langle x_j|x_j\rangle^2 + \sum_{i\neq j} |\langle x_i|x_j\rangle|^2. \tag{1.11}$$

So

$$0 = \sum_{i\neq j} |\langle x_i|x_j\rangle|^2. \tag{1.12}$$

So each inner product must be 0. Thus $\{|x_i\rangle\}$ will be an orthonormal set. We can write any vector in terms of this set, using

$$|y\rangle = I\,|y\rangle = \sum_i |x_i\rangle \langle x_i|y_i\rangle \equiv \sum_i \alpha_i\,|x_i\rangle\,. \tag{1.13}$$

So the set spans the space. Finally, if there exist some coefficients, α_i such that $\sum_i \alpha_i\,|x_i\rangle = 0$, we have

$$\langle x_j| \sum_i \alpha_i\,|x_i\rangle = \alpha_j = \langle x_j|\,0 = 0,\ \forall j. \tag{1.14}$$

So all of the coefficients are 0. Thus the vectors in the set are linearly independent. Taking all of these results together, $\{|x_i\rangle\}$ forms a basis. ■

Equation 1.10 can be used to make a number of proofs very clean, by inserting it into different locations. We can see some examples of this, looking at the idea of the trace of a matrix:

Definition 1 (Trace)
The trace of a matrix, A, is defined to be

$$\mathrm{Tr}(A) = \sum_i \langle i|A|i\rangle \in \mathbb{C}, \tag{1.15}$$

where $\{|i\rangle\}$ is the canonical basis.

The following properties of the trace can be quickly proven:

1. $\mathrm{Tr}(A\,|y\rangle\,\langle x|) = \langle x|A|y\rangle$

   Proof.

   $$\mathrm{Tr}(A\,|y\rangle\,\langle x|) = \sum_i \langle i|A|y\rangle\,\langle x|i\rangle = \langle x|\sum_i |i\rangle\ \langle i|A|y\rangle = \langle x|A|y\rangle$$

   . ■

2. $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$

   Proof.

   $$\mathrm{Tr}(AB) = \sum_i \langle i|AB|i\rangle = \sum_{ij} \langle i|A|j\rangle\ \langle j|B|i\rangle = \sum_{ij} \langle j|B|i\rangle\ \langle i|A|j\rangle$$
   $$= \sum_j \langle j|BA|j\rangle = \mathrm{Tr}(BA).$$

   ■

3. $\mathrm{Tr}(\sum_i \alpha_i A_i) = \sum_i \alpha_i \mathrm{Tr}(A_i)$

   Proof.

   $$\mathrm{Tr}(\sum_i \alpha_i A_i) = \sum_j \langle j|\sum_i \alpha_i A_i|j\rangle = \sum_i \alpha_i \sum_j \langle j|A_i|j\rangle = \sum_i \alpha_i \mathrm{Tr}(A_i).$$

   ■

Now let $\{|\phi_i\rangle\}$ be any orthonormal basis of vectors.

$$\begin{aligned}\sum_i \langle \phi_i|A|\phi_i\rangle &= \mathrm{Tr}(\sum_i A\,|\phi_i\rangle\,\langle\phi_i|)\\ &= \mathrm{Tr}(AI)\\ &= \mathrm{Tr}(A).\end{aligned} \tag{1.16}$$

Thus the trace is basis independent. We should note that the trace of a matrix is also equal to the sum of its eigenvalues: $\mathrm{Tr}(A) = \sum_i \lambda_{A,i}$.

We point out three types of matrices which are of particular interest to us.

Definition 2 (3 Types of Matrices)
A matrix, H, such that $H^\dagger = H$ is called ***Hermitian****.*
A matrix, U, such that $UU^\dagger = U^\dagger U = I$ is called ***unitary****.*
A matrix, A, such that $A^\dagger A = AA^\dagger$ is called ***normal****.*

Note: Both unitary and Hermitian matrices are also normal matrices.

Unitary matrices correspond to the operations through which we manipulate our states. Consider equation 1.3. Taking $|\psi\rangle = |m\rangle + i\,|n\rangle$, where $|m\rangle \neq |n\rangle$ are elementary vectors, we have

$$(\langle m| - i\,\langle n|)A^\dagger A(|m\rangle + i\,|n\rangle) = (\langle m| - i\,\langle n|)(|m\rangle + i\,|n\rangle) = 2. \tag{1.17}$$

But this is also equal to

$$\begin{aligned} &\langle m|A^\dagger A|m\rangle + \langle n|A^\dagger A|n\rangle + i(\langle m|A^\dagger A|n\rangle - \langle n|A^\dagger A|m\rangle) \\ &= \langle m|m\rangle + \langle n|n\rangle + i(\langle m|A^\dagger A|n\rangle - \langle n|A^\dagger A|m\rangle) \\ &= 2 + i(\langle m|A^\dagger A|n\rangle - \langle n|A^\dagger A|m\rangle). \end{aligned} \tag{1.18}$$

Thus $\langle m|A^\dagger A|n\rangle = \langle n|A^\dagger A|m\rangle$. But

$$2 = (\langle m| + \langle n|)A^\dagger A(|m\rangle + |n\rangle) = 2 + \langle m|A^\dagger A|n\rangle + \langle n|A^\dagger A|m\rangle\,. \tag{1.19}$$

So $\langle m|A^\dagger A|n\rangle = -\langle n|A^\dagger A|m\rangle$. Thus the off-diagonal terms of $A^\dagger A$ are 0. And of course the diagonal terms are $\langle m|A^\dagger A|m\rangle = 1$. So $A^\dagger A = I$. This, in turn, implies $AA^\dagger = I$, since $A^\dagger A = I$ implies $\det(A)\det\left(A^\dagger\right) = \det(I) = 1$. So $\det(A) \neq 0$. Thus A has an inverse and so

$$A(A^\dagger A) = A \implies AA^\dagger AA^{-1} = AA^\dagger = I. \tag{1.20}$$

On the other hand, unitary matrices will preserve the norms of all vectors: $\langle x|U^\dagger U|x\rangle = \langle x|I|x\rangle = \langle x|x\rangle$. Thus the manipulations we perform on our states are precisely the unitary operations. We tend to call these unitary manipulations **quantum gates** when we intentionally apply them in a quantum computing scheme. However, they can also be applied unintentionally, by noise. These noisy processes will occupy our attention in chapters 2 and 3.

Normal matrices are, equivalently, those matrices which can be unitarily diagonalized. For this reason, normal matrices are also known as **unitarily diagonalizable matrices**.

Theorem 1.3.2 (Spectral Theorem)
A matrix, A, is normal iff it can be written in the form

$$A = UDU^\dagger, \tag{1.21}$$

with D, a diagonal matrix whose diagonal entries are equal to the eigenvalues of A, and U, a unitary matrix with columns equal to the corresponding eigenvectors.

On the other hand, for any unitary, U, and diagonal matrix, D, $UDU^\dagger$ has eigenvalues equal to the diagonal elements of D, and eigenvectors equal to the columns of U.

We can alternatively write the spectral decomposition using the eigenvalues and eigenvectors of A:

$$A = \sum_{ij} U\left|i\right\rangle \left\langle i\right|D\left|j\right\rangle \left\langle j\right|U^\dagger = \sum_i \lambda_i \left|\hat{u}_i\right\rangle \left\langle \hat{u}_i\right|. \tag{1.22}$$

Note: We write $|\hat{u}\rangle$ to signify that the columns of the unitary matrix are normalized. In fact, the columns of a unitary matrix are orthonormal:

$$\langle \hat{u}_i|\hat{u}_j\rangle = \langle i|UU^\dagger|j\rangle = \delta_{ij}. \tag{1.23}$$

Furthermore, by taking $D = I$, we have

$$I = UIU^\dagger = \sum_i \left|\hat{u}_i\right\rangle \left\langle \hat{u}_i\right|. \tag{1.24}$$

Thus from theorem 1.3.1, we know that the columns of U (the eigenvectors of A) form an orthonormal basis.

The spectral decomposition is a very useful tool in linear algebraic investigations. Unfortunately, not every matrix is normal. To get around this, for any matrix, A, we can consider the Hermitian matrices $AA^\dagger$ and $A^\dagger A$. Performing the spectral decomposition on these matrices gives us

$$\begin{aligned} AA^\dagger &= UD_1U^\dagger \\ A^\dagger A &= VD_2V^\dagger. \end{aligned} \tag{1.25}$$

It can be shown that the eigenvalues of $AA^\dagger$ and $A^\dagger A$ are the same. The two decompositions in equation 1.25 can be "pulled apart" to perform a decomposition of A called the **singular value decomposition**:

Theorem 1.3.3 (Singular Value Decomposition)
Let A be an $n \times m$ matrix. There exist unitary matrices, U $(n \times n)$, V $(m \times m)$, and a rectangular diagonal matrix, Σ, such that

$$A = U\Sigma V^\dagger. \tag{1.26}$$

*The diagonal elements of Σ are the **singular values** of A. The set of the singular values of A is denoted $\sigma(A) \equiv \{\sigma_i\}$. For any A, we have*

$$\sigma_i \in \mathbb{R}_{\geq 0}\ , \quad \forall i. \tag{1.27}$$

Since $A^\dagger A = V\Sigma U^\dagger U\Sigma^T V^\dagger = V\Sigma\Sigma^T V^\dagger$ and $\Sigma\Sigma^T$ is diagonal, with non-zero entries equal to the diagonal entries of Σ, squared, by theorem 1.3.2 the singular values of A are the positive square roots of the eigenvalues of $A^\dagger A$. This leads us to the idea of the *square root* of a matrix.

Definition 3 (Square Root of a Matrix)
*Given a diagonal matrix, D, we define the **(positive) square root** of D to be a diagonal matrix, whose entries are the (positive) square roots of the entries of D:*

$$D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \rightarrow \sqrt{D} = \begin{pmatrix} \sqrt{d_1} & & \\ & \ddots & \\ & & \sqrt{d_n} \end{pmatrix}. \tag{1.28}$$

More generally, given a normal matrix, A, with spectral decomposition $A = UDU^\dagger$, we define its square root to be:

$$\sqrt{A} = U\sqrt{D}U^\dagger. \tag{1.29}$$

Thus the eigenvalues of $\sqrt{A^\dagger A}$ are the singular values of A. We give this matrix a special notation:

Definition 4
*Given a matrix, A, the **absolute value** of that matrix is*

$$|A| \equiv \sqrt{A^\dagger A}. \tag{1.30}$$

We see that for any Hermitian matrix, $H = UDU^\dagger$, we have $|H| = \sqrt{H^2} = U\sqrt{D^2}U^\dagger$, and so the singular values are the absolute values of the eigenvalues (noting that the

eigenvalues of a Hermitian matrix are real, since $\lambda = \langle x|H|x\rangle = \langle x|H^\dagger|x\rangle = \langle x|H|x\rangle^* = \lambda^*$, using normalized eigenvector, $|x\rangle$). For any Hermitian matrix whose eigenvalues are all non-negative, the singular values and eigenvalues match.

Definition 5 (Positive Semi-Definite Matrix)
*A matrix is called **positive semi-definite** (PSD) if it is Hermitian and all of its eigenvalues are non-negative. We denote this by $A \geq 0$.*

For any A, $|A|$ is PSD. If A is PSD, then $|A| = A$. The absolute value has one other major appearance that we will be interested in — the **polar decomposition**:

Theorem 1.3.4 (Polar Decomposition of a Matrix)
Every matrix, A, can be written as

$$A = UP, \tag{1.31}$$

where U is a unitary matrix, and $P = |A|$ is PSD.

This breakdown of matrices is similar to the polar decomposition of complex numbers: $z = |z|e^{i\phi}$. Such a decomposition allows us to separate two kinds of behaviour of complex numbers – angular and radial. We will return to this idea in chapter 3, when we develop a polar decomposition of quantum channels, separating coherent and decoherent behaviour.

Up until now, we have focused on viewing quantum states as vectors, $|\psi\rangle$. It is often more useful to view them as matrices, as it elevates us into a new regime, wherein we develop the faculties to tackle new questions.

1.4 Density Matrices

In physical applications, there is always a level of uncertainty as to which state our system is in. Perhaps we had meant to prepare our system in an initial state, $|\psi\rangle$, but due to inaccuraries in the experimental setup, we had some probability, p_2 of preparing the system in the state $|\phi\rangle$. We would express the state of our system as a **density matrix**:

$$\rho = (1 - p_2)\,|\psi\rangle\langle\psi| + p_2\,|\phi\rangle\langle\phi|\,. \tag{1.32}$$

In fact, we *define* density matrices to be of this form:

Definition 6 (Density Matrix)
*A **density matrix** is any matrix of the form*

$$\rho = p_1 \left|\psi_1\right\rangle \left\langle\psi_1\right| + \ldots + p_m \left|\psi_m\right\rangle \left\langle\psi_m\right|, \tag{1.33}$$

where $m \in \mathbb{Z}$, $\{|\psi_i\rangle\}$ are quantum states, and $\{p_i\}$ are probabilities: i.e. $p_i \in \mathbb{R}_{\geq 0}$, $\forall i$ and $\sum_i p_i = 1$.

Since the p_i in the above definition are real, and $(|\psi\rangle\langle\psi|)^\dagger = |\psi\rangle\langle\psi|$, $\forall\, |\psi\rangle$, it immediately follows that all density matrices are Hermitian: $\rho^\dagger = \rho$. Furthermore, we can see that $\mathrm{Tr}(\rho) = \sum_i p_i \langle\psi_i|\psi_i\rangle = \sum_i p_i = 1$, since states are normalized. Finally, consider any eigenvalue, λ, of a density matrix, with normalized eigenvector $|x\rangle$. We have

$$\begin{aligned} \lambda &= \langle x|\rho|x\rangle \\ &= \langle x|(p_1 \left|\psi_1\right\rangle \left\langle\psi_1\right| + \ldots + p_m \left|\psi_m\right\rangle \left\langle\psi_m\right|)|x\rangle \\ &= \sum_i p_i |\langle x|\psi_i\rangle|^2 \\ &\geq 0. \end{aligned} \tag{1.34}$$

So every eigenvalue of a density matrix is non-negative. Thus density matrix are positive semi-definite (since they are Hermitian). Density matrices are sometimes defined by the above three properties:

1. $\rho^\dagger = \rho$.
2. $\mathrm{Tr}(\rho) = 1$.
3. $\lambda_i \geq 0$, $\forall i$.

Any matrix which satisfies these properties can be spectrally decomposed, and so would be of the form

$$\rho = \sum_i \lambda_i \left|x_i\right\rangle \left\langle x_i\right|, \tag{1.35}$$

where $\sum_i \lambda_i = Tr(\rho) = 1$, and $\{|x_i\rangle\}$ are orthonormal, giving us something satisfying definition 6. So we see that the formulations are equivalent.

In both the spectral decomposition and definition 6, density matrices have the form $\sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$. The only difference is whether or not the $\{|\psi_i\rangle\}$ are orthonormal. For the

sake of ease, whenever we want to expand a density matrix in terms of state vectors, we will use the spectral decomposition.

Density matrices in which all of the probabilities are 0 except for one – $\rho = |\psi\rangle \langle\psi|$ – are called **pure states**. The general form of a density matrix, à la equation 1.33, is called a **mixed state**. We might suspect that at the opposite end of the spectrum from a pure state, is one in which all events have equal probability. Recalling that a sum over any orthonormal basis brings us to the identity, we arrive at

$$\rho = \sum_i \frac{1}{d} |\psi_i\rangle \langle\psi_i| = \frac{1}{d} I. \tag{1.36}$$

This is known as a **maximally mixed state**. Recall that we can manipulate quantum states through operators: $U|\psi\rangle$. The appropriate promotion of this concept to the density matrix formalism is unitary conjugation: $U\rho U^\dagger = \sum_i p_i U |\psi_i\rangle \langle\psi_i| U^\dagger$. In the next chapter we will develop a method for expressing the idea that there can be uncertainty in which operation we are performing on a state, as well. Just as density matrices encapsulate the idea of not being certain as to which pure state a quantum mechanical system is in, quantum channels will make rigorous the idea of being unsure as to which operation we are applying to our quantum system.

1.5 Tensor Product

Another way to elevate the quantum franchise we have developed thus far is to introduce a way to stitch different quantum systems together. If we have two quantum systems, $\mathcal{H}_A$ and $\mathcal{H}_B$, we can study them independently, but we would also like a way to allow them to interact. What we grab is the **tensor product**, $\mathcal{H}_A \otimes \mathcal{H}_B$. A common way that tensor products appear in quantum computing is in the coupling of the environment to a system of interest: the Hilbert space of the total system would be $\mathcal{H}_{\text{environment}} \otimes \mathcal{H}_{\text{system}}$.

The elements of a tensor space are of the form $\sum_{ij} a_{ij} |x_i\rangle \otimes |y_j\rangle$, where $|x_i\rangle \in \mathcal{H}_A$ and $|y_j\rangle \in \mathcal{H}_B \ \forall i, j$. We can also form tensor products of matrix spaces: $\mathbb{M}_{d_1}(\mathbb{C}) \otimes \mathbb{M}_{d_2}(\mathbb{C})$. If the eigenvalues of the matrices A and B are $\{\lambda_{A,i}\}_{i=1}^m$ and $\{\lambda_{B,j}\}_{j=1}^n$, with corresponding eigenvectors $\{|x_{A,i}\rangle\}$, and $\{|x_{B,j}\rangle\}$, the eigenvalues of the tensor product, $A \otimes B$, are $\{\lambda_{A,i}\, \lambda_{B,j}\}_{i=1,j=1}^{i=m,j=n}$, with eigenvectors $\{|x_{A,i}\rangle \otimes |x_{B,j}\rangle\}$.

Viewing eigenvalues as probabilities and states as events, it follows that $\lambda_{A,i}\, \lambda_{B,j}$ describes the probability of event $|x_{A,i}\rangle \otimes |x_{B,j}\rangle$. Then $\rho_A \otimes \rho_B$ must describe two independent

random variables. Or, from the perspective of quantum physics, two independent (non-interacting) quantum states (since the probabilities factor). In general, elements of the tensor space will be of the form $\sum_{ijkl} a_{ijkl} E_{ij} \otimes E_{kl}$, and thus will describe interacting systems.

We list a quick result about tensor products before moving on. Using the fact that the eigenvalues of $A \otimes B$, are $\{\lambda_{A,i}\, \lambda_{B,j}\}$, we have

$$\begin{aligned} \mathrm{Tr}(A \otimes B) &= \sum_{ij} \lambda_{A,i}\, \lambda_{B,j} \\ &= \sum_{i} \lambda_{A,i} \sum_{j} \lambda_{B,j} \\ &= \mathrm{Tr}(A)\mathrm{Tr}(B). \end{aligned} \tag{1.37}$$

We can also define the partial trace of a tensor product:

Definition 7 (Partial Trace)
Given a tensor product, $A_1 \otimes \ldots \otimes A_n$, the ***i-th partial trace*** *is*

$$\mathrm{Tr}_i(A_1 \otimes \ldots \otimes A_n) = \mathrm{Tr}(A_i) A_1 \otimes \ldots \otimes A_{i-1} \otimes A_{i+1} \otimes \ldots \otimes A_n. \tag{1.38}$$

We quickly make note of the rules of tensor manipulations:

$$\alpha(|x\rangle \otimes |y\rangle) = (\alpha\, |x\rangle) \otimes |y\rangle = |x\rangle \otimes (\alpha\, |y\rangle). \tag{1.39a}$$

$$(A \otimes B)(|x\rangle \otimes |y\rangle) = A\, |x\rangle \otimes B\, |y\rangle\,. \tag{1.39b}$$

$$\alpha(A \otimes B) = (\alpha A) \otimes B = A \otimes (\alpha B). \tag{1.39c}$$

$$A \otimes (B + C) = A \otimes B + A \otimes C. \tag{1.39d}$$

$$(A + B) \otimes C = A \otimes C + B \otimes C \tag{1.39e}$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \tag{1.39f}$$

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger. \tag{1.39g}$$

For all $\alpha \in \mathbb{C}$, $|x_i\rangle \in \mathcal{H}^{d_1}$ and $|y_j\rangle \in \mathcal{H}^{d_2}$, and matrices A, B, C in the appropriate $\mathbb{M}_n(\mathbb{C})$.

1.6 Vectorization

We shall find as we move forward, that we can get a lot of mileage out of the idea of transforming a matrix into a vector, by stacking its columns one on top of the other:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}. \tag{1.40}$$

We formalize this with the following definition.

Definition 8 (Vectorization)
Write a given matrix, A, in terms of the canonical basis: $A = \sum_{ij} a_{ij} |i\rangle \langle j|$. Then the ***vectorization*** *of this matrix is:*

$$|A\rangle = \sum_{ij} a_{ij} |j\rangle \otimes |i\rangle . \tag{1.41}$$

In particular, $|i\rangle \langle j| \to |j\rangle \otimes |i\rangle$. Given a outer product of vectors, $|x\rangle \langle y|$ we have:

$$\begin{aligned} |x\rangle \langle y| &= \sum_{ij} x_i y_j |i\rangle \langle j| \\ \text{Vectorize} \to &\sum_{ij} x_i y_j |j\rangle \otimes |i\rangle \\ &= \sum_j y_j |j\rangle \otimes \sum_i x_i |i\rangle \\ &= |y\rangle \otimes |x\rangle . \end{aligned} \tag{1.42}$$

With the above relation and the following two lemmas, we shall be able to take the vectorization of any matrix, with ease.

Lemma 1.6.1
Vectorization is a linear map.

Proof.

$$\begin{aligned} |\alpha A + \beta B\rangle &= \sum_{ij} (\alpha A + \beta B)_{ij} |j\rangle \otimes |i\rangle \\ &= \sum_{ij} \alpha (A)_{ij} |j\rangle \otimes |i\rangle + \sum_{ij} \beta (B)_{ij} |j\rangle \otimes |i\rangle \\ &= \alpha |A\rangle + \beta |B\rangle . \end{aligned} \tag{1.43}$$

■

Lemma 1.6.2
Given three matrices, A, B, C, the vectorization of the product, ABC, is

$$|ABC\rangle = (C^T \otimes A)\,|B\rangle\,. \tag{1.44}$$

Proof.

$$\begin{aligned}
|ABC\rangle &= \sum_i \sum_j (ABC)_{ij}\,|j\rangle \otimes |i\rangle \\
&= \sum_{ijkl} \langle i|\,A\,|k\rangle\,\langle k|\,B\,|l\rangle\,\langle l|\,C\,|j\rangle\,|j\rangle \otimes |i\rangle \\
&= \sum_{ijkl} \langle k|\,B\,|l\rangle\,|j\rangle\,\langle j|\,C^T\,|l\rangle \otimes |i\rangle\,\langle i|\,A\,|k\rangle \\
&= \sum_{kl} B_{kl} C^T\,|l\rangle \otimes A\,|k\rangle \\
&= (C^T \otimes A)\,|B\rangle
\end{aligned} \tag{1.45}$$

■

Recall that the concept of normalization for density matrices ($\mathrm{Tr}(\rho) = 1$) differs from that of states ($\langle\psi|\psi\rangle = 1$). In transforming from a matrix to a vector, we may find that we need to renormalize:

$$A \to \frac{1}{\sqrt{\langle A|A\rangle}}\,|A\rangle\,. \tag{1.46}$$

For instance, taking the maximally mixed state,

$$\frac{1}{d} I = \frac{1}{d} \sum_i |i\rangle\,\langle i|\,, \tag{1.47}$$

and performing vectorization, leads to an un-normalized state, as $\langle I/d|I/d\rangle = 1/d \neq 1$. The properly normalized state,

$$\frac{1}{\sqrt{d}}\,|I\rangle = \frac{1}{\sqrt{d}} \sum_i |i\rangle \otimes |i\rangle\,, \tag{1.48}$$

is known as a **maximally entangled state**. It will come into play in chapter 2.

Translating from matrices to vectors also reveals a natural inner product on the space of matrices:

$$\begin{aligned}\langle A|B\rangle &= \sum_{ijkl} a_{ij}^* b_{kl} (\langle j| \otimes \langle i|)(|l\rangle \otimes |k\rangle) \\ &= \sum_{ij} a_{ij}^* b_{ij} = \sum_{ij} \langle j| A^\dagger |i\rangle \langle i| B |j\rangle \\ &= \mathrm{Tr}(A^\dagger B).\end{aligned} \tag{1.49}$$

We call $\mathrm{Tr}(A^\dagger B)$ the **Hilbert-Schmidt inner product** of A and B. This inner product provides us a simple way of formulating orthonormal bases of matrices, which will be useful in the next chapter. We define the **Frobenius norm** to be $\|A\|_F \equiv \sqrt{\langle A|A\rangle}$.

Chapter 2

Quantum Channels

2.1 Theoretical Framework

In chapter 1, we considered manipulations of quantum states: $A|x\rangle$. We imposed the restriction $\langle x|A^\dagger A|x\rangle = \langle x|x\rangle \ \forall |x\rangle$, and arrived at the set of unitary matrices. We then promoted states to density matrices, to allow us to consider uncertainty in our states. The action of unitary matrices became conjugation, and we found that we could manipulate our states as $U\rho U^\dagger$. However, the set of maps which preserve density matrices contains much more than just these conjugations. The general maps we will be concerned with are called **quantum channels**, and they will allow us to expand our range of investigation. We will be free to study situations in which we are not completely certain as to how our states are being manipulated, and we will be afforded a perfect regime for studying noise.

The natural way to view quantum channels is as linear maps from the set of quantum states to itself. It will turn out to be easier to extend our scope a bit, however, and formulate channels as maps from the set of $d \times d$ complex matrices to itself, instead: $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$.

Recall, in section 1.4 we listed three properties defining density matrices. If we are to preserve the set of quantum states by our channels, then those defining properties must be preserved. This leads to the following constraints:

1. Trace preservation: $\mathrm{Tr}(\mathcal{A}(\rho)) = \mathrm{Tr}(\rho) \ \forall \rho \in \mathbb{M}_d(\mathbb{C})$.
2. Positivity preservation: $\mathcal{A}(\rho)$ is positive semi-definite $\forall \rho \geq 0$.

The preservation of positivity turns out to not be a strong enough requirement for quantum channels, when we consider physical implications. Consider a tensor space, $\mathbb{M}_{\text{env}}(\mathbb{C}) \otimes \mathbb{M}_{\text{sys}}(\mathbb{C})$. It is completely possible for our channel to only act on $\mathbb{M}_{\text{sys}}(\mathbb{C})$, and leave the other subsystem alone: $I \otimes \mathcal{A}$. In such a situation, we should still map density matrices to density matrices. Thus we make the following requirement:

Definition 9 (K-Positivity)
*A map, $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$, is called **positive** if $\mathcal{A}(M) \geq 0, \forall M \geq 0$.*
*Let I_k denote the $k \times k$ identity matrix. $\mathcal{A}$ is **k-positive** if $I_k \otimes \mathcal{A}$ is positive. A map is **completely positive** if it is k-positive $\forall k$.*

By the physical argument above, all quantum channels must be completely positive.

Note: If we have a quantum channel on a tensor space, $\mathcal{A} : \mathbb{M}_{\text{env}}(\mathbb{C}) \otimes \mathbb{M}_{\text{sys}}(\mathbb{C}) \to \mathbb{M}_{\text{env}}(\mathbb{C}) \otimes \mathbb{M}_{\text{sys}}(\mathbb{C})$, we can define a "reduced map" on one of the tensor factors, say $\mathcal{A}|_{\text{sys}} : \mathbb{M}_{\text{sys}}(\mathbb{C}) \to \mathbb{M}_{\text{sys}}(\mathbb{C})$, as follows: For any matrix, $\rho_S \in \mathbb{M}_{\text{sys}}(\mathbb{C})$, there exists a matrix, $\rho_{ES} \in \mathbb{M}_{\text{env}}(\mathbb{C}) \otimes \mathbb{M}_{\text{sys}}(\mathbb{C})$, such that $\text{Tr}_1(\rho_{ES}) = \rho_S$. In fact, such a ρ_{ES} is not unique: $E_{ii} \otimes \rho_S$ works for any i, for example. We can define maps,

$$\begin{aligned} \mathcal{A}|_{\text{sys}} : \mathcal{H}_{\text{sys}} &\to \mathcal{H}_{\text{sys}} \\ \mathcal{A}|_{\text{sys}}(\rho_S) &\equiv \text{Tr}_1(\mathcal{A}(\rho_{ES})), \ \forall \rho_S \in \mathcal{H}_{\text{sys}}, \end{aligned} \tag{2.1}$$

by selecting ρ_{ES} for each ρ_S.

It is *not at all* assured that $\mathcal{A}|_{\text{sys}}$ will be completely positive for a given choice of ρ_{ES}. Positive but not completely positive maps arise on these reduced systems when sufficient correlations are present between the sytem and environment [5]. Such positive, not completely positive, reduced maps can be considered physical manipulations on the system, and are well worth studying. However, they are not considered quantum channels, and we will not consider them here. By definition, quantum channels must be completely positive.

Complete positivity implies 1-positivity, which means $1 \otimes \mathcal{A} = \mathcal{A}$ is a positive map. So we can drop the positivity condition, in place of complete positivity.

Note: If $\mathcal{A}$ is trace preserving, then so is $I_n \otimes \mathcal{A}$, $\forall n$. For if $A = \sum_{ijkl} a_{ijkl} E_{ij} \otimes E_{kl}$ is a

general state in $\mathbb{M}_n(\mathbb{C}) \otimes \mathbb{M}_d(\mathbb{C})$, then

$$\begin{aligned} \mathrm{Tr}(I_n \otimes \mathcal{A}(\sum_{ijkl} a_{ijkl} E_{ij} \otimes E_{kl})) &= Tr(\sum_{ijkl} a_{ijkl} E_{ij} \otimes \mathcal{A}(E_{kl})) \\ &= \sum_{ijkl} a_{ijkl} \mathrm{Tr}(E_{ij}) \mathrm{Tr}(\mathcal{A}(E_{kl})) \\ &= \sum_{ijkl} a_{ijkl} \mathrm{Tr}(E_{ij}) \mathrm{Tr}(E_{kl}) \\ &= \mathrm{Tr}(A). \end{aligned} \tag{2.2}$$

So there are no additional conditions required with respect to trace preservation, when considering tensor spaces. Thus the appropriate set of conditions, defining the set of quantum channels, is the following:

Definition 10 (Quantum Channels)
A ***quantum channel*** *is a linear map,* $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$ *that is completely positive and trace preserving. Such maps are also called* ***CPTP maps****.*

We can now begin to explore the set of quantum channels. Unfortunately, while the CPTP condition is intuitive, it provides few footholds for investigation. By representing our quantum channels by matrices, or sets of matrices, we will find them easier to study.

2.2 Representations of Quantum Channels

There are many ways of expressing quantum channels in terms of matrices. We shall review three here, which will prove most useful to us: the Kraus representation, the Choi matrix, and the Liouville representation.

2.2.1 Kraus representation

Perhaps the most common way of expressing the action of a quantum channel is through the **Kraus representation**:

Theorem 2.2.1 (Kraus Representation)
A map $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$ *is completely positive iff there exists a set of matrices* $\{A_j\}$ *(the* ***Kraus operators*** *of* $\mathcal{A}$*) such that*

$$\mathcal{A}(\rho) = \sum_j A_j \rho A_j^\dagger \ , \ \forall \rho \in \mathbb{M}_d(\mathbb{C}). \tag{2.3}$$

Furthermore, $\mathcal{A}$ *is trace preserving iff*

$$\sum_j A_j^\dagger A_j = I. \tag{2.4}$$

Proof. See [27]. ■

From trace preservation, quantum channels with only a single Kraus operator are unitary conjugations, $U\rho U^\dagger$, in which the system is manipulated in a reversible fashion. The actions of quantum channels with more than one Kraus operator are not completely reversible.

If $\mathcal{A}$ and $\mathcal{B}$ have Kraus operators $\{A_i\}$ and $\{B_j\}$ then the composition $\mathcal{AB}$ has Kraus operators $\{A_i B_j\}$:

$$\mathcal{AB}(\rho) = \mathcal{A}(\mathcal{B}(\rho)) = \sum_i \sum_j A_i B_j \rho B_j^\dagger A_i^\dagger. \tag{2.5}$$

This is a simple way of producing Kraus operators for compositions, but the size of the set of Kraus operators it produces grows quickly. There is often a smaller set of Kraus operators available (it can be shown that for any channel $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$ there exists a set of Kraus operators of size $\leq d^2$). The set of Kraus operators for a given quantum channel is not unique — given a set of Kraus operators, $\{A_j\}$, for a channel, $\mathcal{A}$, we can create a new set of Kraus operators, describing the same channel, from an appropriate linear combination of the $\{A_j\}$. On the other hand, every set of Kraus operators for the channel $\mathcal{A}$ can be written in this way:

Theorem 2.2.2 (Kraus Freedom)
Let $\{A_i\}_{i=1}^m$ and $\{B_i\}_{j=1}^n$ be sets of Kraus operators for quantum channels $\mathcal{A}$ and $\mathcal{B}$, respectively. Extend the smaller set with zero-matrices, so that the sets have the same size. Then $\mathcal{A} = \mathcal{B}$ iff there exists a $\max\{m,n\} \times \max\{m,n\}$ unitary matrix, U, with entries u_{jk}, such that

$$A_j = \sum_k u_{jk} B_k. \tag{2.6}$$

Proof. See [27]. ■

We see then that there is a degree of freedom available in our choice of Kraus operators. In chapter 3, we will repeatedly use a particular choice of Kraus operators, known as the **canonical Kraus operators**, for every quantum channel we encounter. In order to arrive at these Kraus operators, we introduce the Choi matrix.

2.2.2 Choi Matrix

Definition 11 (Choi Matrix)
*The **Choi matrix** of a quantum channel, $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$, is defined to be the $d^2 \times d^2$ matrix,*

$$\beth_{\mathcal{A}} \equiv I \otimes \mathcal{A}(|I\rangle \langle I|) = \sum_{ij} E_{ij} \otimes \mathcal{A}(E_{ij}). \tag{2.7}$$

It may not be obvious from the definition, but the Choi matrix does a fantastic job of capturing all of the information about a quantum channel that we are interested in (and more!). Consider the following theorem:

Theorem 2.2.3 (Choi Characterization)
Let $\mathcal{A} : \mathbb{M}_d(\mathbb{C}) \to \mathbb{M}_d(\mathbb{C})$ be a linear map. The Choi matrix, $\beth_{\mathcal{A}}$ is positive semi-definite iff $\mathcal{A}$ is completely positive, and $Tr_2(\beth_{\mathcal{A}}) = I$ iff $\mathcal{A}$ is trace preserving.

Proof. We prove the trace preservation part of the theorem and refer to [6] for the complete

positivity.

$$\begin{aligned} Tr_2(\beth_{\mathcal{A}}) &= \sum_{ij} E_{ij} \otimes \mathrm{Tr}(\mathcal{A}(E_{ij})) \\ &= \sum_{ijkn} |i\rangle \langle j| \langle n| A_k |i\rangle \langle j| A_k^\dagger |n\rangle \\ &= \sum_{ijkn} |i\rangle \langle i| A_k^T |n\rangle \langle n| A_k^* |j\rangle \langle j| \\ &= (\sum_k A_k^\dagger A_k)^T \end{aligned} \tag{2.8}$$

By theorem 2.2.1 $\mathcal{A}$ is trace preserving iff $Tr_2(\beth_{\mathcal{A}})^T = \sum_k A_k^\dagger A_k = I$ which is iff $Tr_2(\beth_{\mathcal{A}}) = I$. ■

Since the Choi matrix is PSD for any quantum channel, we can perform a spectral decomposition, $\beth_{\mathcal{A}} = \sum_j \lambda_j |\hat{a}_j\rangle \langle \hat{a}_j|$. If we unvectorize $\sqrt{\lambda_j} |\hat{a}_j\rangle$, we receive a $d \times d$ matrix, A_j. It turns out that the set of matrices, $\{A_j\}$, is a set of Kraus operators for the channel $\mathcal{A}$ [6]. These are known as the **canonical Kraus operators** for $\mathcal{A}$. Taking the inner product of the Kraus operators, we see

$$\mathrm{Tr}(A_i^\dagger A_j) = \sqrt{\lambda_i \lambda_j} \langle \hat{a}_i | \hat{a}_j \rangle = \delta_{ij} \sqrt{\lambda_i \lambda_j}. \tag{2.9}$$

Thus the canonical Kraus operators are orthogonal and have squared norms equal to the corresponding eigenvalue: $\mathrm{Tr}(A_i^\dagger A_i) = \|A_i\|_F^2 = \lambda_i$. Noting that $\mathrm{Tr}(\beth_{\mathcal{A}}) = \mathrm{Tr}(I) = d$, we can interpret each λ_i/d as the probability of the i-th canonical Kraus operator being applied.

2.2.3 Liouville Representation

While the Choi matrix captures many characteristics of quantum channels, and provides the groundwork for the theory we develop in chapter 3, it fails at recognizing the composition of channels in an intuitive way: $\beth_{\mathcal{A}} \beth_{\mathcal{B}} \neq \beth_{\mathcal{AB}}$. We would like to have a channel representation which extends channel composition to matrix multiplication. This is the representation we create here.

Let $\{M_i\}$ be a basis of $n \times n$ matrices which is orthonormal under the Hilbert-Schmidt inner product. For instance, we could take the set of elementary matrices, $\{E_{ij}\}$:

$$\langle E_{ij} | E_{kl} \rangle = \mathrm{Tr}(E_{ij}^\dagger E_{kl}) = \mathrm{Tr}(|j\rangle \langle i|k\rangle \langle l|) = \delta_{ik} \delta_{jl}. \tag{2.10}$$

We will write our intended representation of the channel $\mathcal{A}$, as $\boxed{\mathcal{A}}$. It is called the **Liouville representation**. We will define our representation through a very simple requirement:

$$\boxed{\mathcal{A}}\,|M\rangle = |\mathcal{A}(M)\rangle\,,\ \forall M \in \mathbb{M}_d(\mathbb{C}). \tag{2.11}$$

With this requirement in place, it is easy to see that the product of the Liouville representations of two channels is just the representation of the composition of the channels:

$$\begin{aligned}
\boxed{\mathcal{A}}\boxed{\mathcal{B}} &= \sum_j \boxed{\mathcal{A}}\boxed{\mathcal{B}}\,|M_j\rangle\langle M_j| \\
&= \sum_j \boxed{\mathcal{A}}\,|\mathcal{B}(M_j\rangle)\langle M_j| \\
&= \sum_j |\mathcal{A}\mathcal{B}(M_j)\rangle\langle M_j| \\
&= \sum_j \boxed{\mathcal{A}\mathcal{B}}\,|M_j\rangle\langle M_j| \\
&= \boxed{\mathcal{A}\mathcal{B}}.
\end{aligned} \tag{2.12}$$

Where in the first and final line we used the fact that $\sum_j |M_j\rangle\langle M_j| = I$ since $\{M_i\}$ (and thus $\{|M_i\rangle\}$) is an orthonormal basis.

Thus we arrive at a representation which neatly captures the action of quantum channels. If we enforce equation 2.11 on an orthonormal basis of matrices, we can extend linearly so that 2.11 holds for all vectorized matrices.

If $\boxed{\mathcal{A}}\,|M_i\rangle = |\mathcal{A}(M_i)\rangle$ for an orthonormal basis of matrices, $\{M_i\}$, then

$$\boxed{\mathcal{A}}\,|\rho\rangle = \boxed{\mathcal{A}}\left|\sum_i p_i M_i\right\rangle = \sum_i p_i\,\boxed{\mathcal{A}}\,|M_i\rangle = \left|\sum_i p_i \mathcal{A}(M_i)\right\rangle = |\mathcal{A}(\rho)\rangle\,,\ \forall\rho. \tag{2.13}$$

Assuming there exists a representation, $\boxed{\mathcal{A}}$, such that 2.11 is satisfied for some orthonormal basis of matrices, $\{M_i\}$, we would have:

$$\boxed{\mathcal{A}} = \sum_i \boxed{\mathcal{A}}\,|M_i\rangle\langle M_i| = \sum_i |\mathcal{A}(M_i)\rangle\langle M_i| \tag{2.14}$$

On the other hand, we can see that if we just take this equation to be the definition of the

Liouville representation, then

$$\begin{aligned}\boxed{\mathcal{A}}\,|M\rangle = \sum_i |\mathcal{A}(M_i)\rangle\langle M_i| \sum_j m_j\,|M_j\rangle &= \sum_j m_j\,|\mathcal{A}(M_j)\rangle \\ &= \left|\mathcal{A}(\sum_j m_j M_j)\right\rangle \\ &= |\mathcal{A}(M)\rangle \;\; \forall M \in \mathbb{M}_d(\mathbb{C}).\end{aligned} \tag{2.15}$$

Thus this is an appropriate definition.

Definition 12 (Liouville Representation of a Quantum Channel)
*Let $\{M_i\}$ be a basis of $n \times n$ matrices, orthonormal with respect to the Hilbert-Schmidt inner product. The **Liouville representation** of a quantum channel, $\mathcal{A}$, with respect to this basis is:*

$$\boxed{\mathcal{A}} \equiv \sum_i |\mathcal{A}(M_i)\rangle\langle M_i|\,. \tag{2.16}$$

Note: This representation, $f(\mathcal{A}) = \boxed{\mathcal{A}}$, is well-defined, since if $\mathcal{A} = \mathcal{B}$ then $|\mathcal{A}(M_j)\rangle = |\mathcal{B}(M_j)\rangle\ \forall j$, so the representations will be the same. It is also injective, since if $\boxed{\mathcal{A}} = \boxed{\mathcal{B}}$, then by the defining property of the representation, equation 2.11,

$$\begin{aligned}|\mathcal{A}(\rho)\rangle = \boxed{\mathcal{A}}\,|\rho\rangle = \boxed{\mathcal{B}}\,|\rho\rangle = |\mathcal{B}(\rho)\rangle \\ \implies \mathcal{A}(\rho) = \mathcal{B}(\rho),\ \forall\rho.\end{aligned} \tag{2.17}$$

However, it is not surjective. Consider the vectorizations of the elementary matrices, $\{|E_{ij}\rangle\}$. They form an orthonormal basis, and so we can consider them in definition 12. We can pluck out the first element of $\boxed{\mathcal{A}}$ with

$$\begin{aligned}\langle E_{11}|\boxed{\mathcal{A}}\,|E_{11}\rangle &= \langle E_{11}|\sum_{ij}|\mathcal{A}(E_{ij})\rangle\langle E_{ij}|E_{11}\rangle \\ &= \langle E_{11}|\mathcal{A}(E_{11})\rangle \\ &= \mathrm{Tr}(|1\rangle\langle 1|\sum_k A_k\,|1\rangle\langle 1|\,A_k^\dagger) \\ &= \sum_k |\,\langle 1|A_k|1\rangle|^2\end{aligned} \tag{2.18}$$

Thus the first entry in $\boxed{\mathcal{A}}$, with respect to the elementary basis, must be non-negative, preventing surjectivity. Of course, this is just one choice of basis in definition 12. However, we can show that the Liouville representation of any channel is independent of the basis of matrices, $\{M_i\}$, that we pick:

Let $\boxed{\mathcal{A}}_M$ and $\boxed{\mathcal{A}}_N$ be the representation of $\mathcal{A}$ with respect to two orthonormal bases, $\{M_i\}$ and $\{N_i\}$.

$$\langle E_{ij}|\boxed{\mathcal{A}}_M|E_{kl}\rangle = \langle E_{ij}|\mathcal{A}(E_{kl})\rangle = \langle E_{ij}|\boxed{\mathcal{A}}_N|E_{kl}\rangle \ , \forall i,j,k,l. \tag{2.19}$$

The vectorized elementary matrices pick out each element of the representations. So we must have $\boxed{\mathcal{A}}_M = \boxed{\mathcal{A}}_N$. Because of this, we will consider the simplest orthonormal basis — the elementary matrices — in order arrive at an explicit form for the Liouville representation.

$$\begin{aligned}
\boxed{\mathcal{A}} &= \sum_{ij} |\mathcal{A}(E_{ij})\rangle\langle E_{ij}| = \sum_{ijk} \left|A_k E_{ij} A_k^\dagger\right\rangle\langle E_{ij}| \\
&= \sum_{ijk} A_k^* \otimes A_k |E_{ij}\rangle\langle E_{ij}| \\
&= \sum_{ijk} A_k^* \otimes A_k |j\rangle \otimes |i\rangle (\langle j| \otimes \langle i|) \\
&= \sum_k A_k^* \otimes A_k.
\end{aligned} \tag{2.20}$$

In the above equations, we picked some set of Kraus operators $\{A_k\}$ for the quantum channel, $\mathcal{A}$. But the Liouville representation of a channel will also be independent of whichever set of Kraus operators we choose to represent our channel, since by definition 12, the Liouville representation can be written purely in terms of the action of the channel.

Given a channel, $\mathcal{A}$, with Kraus operators $\{A_i\}$, we define the **conjugate channel**, $\mathcal{A}^\dagger$, to have Kraus operators $\{A_i^\dagger\}$. Then we see

$$\boxed{\mathcal{A}^\dagger} = \sum_k (A_k^\dagger)^* \otimes A_k^\dagger = (\sum_k A_k^* \otimes A_k)^\dagger = \boxed{\mathcal{A}}^\dagger. \tag{2.21}$$

We might be interested in what happens when we perform the Hilbert-Schmidt inner product on the Liouville representation of channels:

$$\begin{aligned}
\left\langle \boxed{\mathcal{A}} \middle| \boxed{\mathcal{B}} \right\rangle = \mathrm{Tr}(\boxed{\mathcal{A}}^\dagger \boxed{\mathcal{B}}) &= \sum_{jk} \mathrm{Tr}\left((A_j^* \otimes A_j)^\dagger (B_k^* \otimes B_k)\right) \\
&= \sum_{jk} \left|\mathrm{Tr}\left(A_j^\dagger B_k\right)\right|^2
\end{aligned} \tag{2.22}$$

We can investigate this inner product from both the perspective of the Liouville representation and the Kraus representation. Moreover, if we write the vectorization of the Liouville representation as:

$$\left|\boxed{\mathcal{A}}\right\rangle = \sum_{ij} |E_{ij}\rangle \otimes |\mathcal{A}(E_{ij})\rangle \tag{2.23}$$

there is a clear relationship to the Choi matrix. We have

$$\begin{aligned}\left\langle\boxed{\mathcal{A}}\middle|\boxed{\mathcal{B}}\right\rangle &= \sum_{ijkl} \langle E_{ij}|E_{kl}\rangle \langle \mathcal{A}(E_{ij})|\mathcal{B}(E_{kl})\rangle \\ &= \sum_{ijkl} \mathrm{Tr}(E_{ij}^\dagger E_{kl} \otimes \mathcal{A}(E_{ij})^\dagger \mathcal{B}(E_{kl})) \\ &= \langle \beth_{\mathcal{A}} | \beth_{\mathcal{B}} \rangle \,.\end{aligned} \tag{2.24}$$

So we see that this object, this inner product of quantum channel representations, appears, regardless of which of the three representations we choose to use. It may then be something worth investigating. We shall see that it will play a key role in what follows. We begin by discussing its relationship to two figures of merit for quantum channels: the fidelity and the unitarity.

2.3 Fidelity and Unitarity

In any given quantum computing algorithm, we expect to be able to perform specific operations on our quantum states. In practice, there is always some probability that we will perform an operation incorrectly, leading to an inappropriate state. We can measure the level of accuracy in our quantum applications using the concept of the **fidelity**. Given two pure states, $|\psi\rangle$ and $|\phi\rangle$, the fidelity between the states is defined to be $|\langle\psi|\phi\rangle|$. The fidelity provides us with information about how closely related two states are, and thus can be used to measure the accuracy of our computations. The fidelity between two pure states can be extended to a measure of fidelity between general states. However, we will be more interested in pursuing a notion of fidelity related to quantum channels. One way of comparing quantum channels is by comparing their outputs upon acting on the same input state. This leads to the concept of **M-fidelities**.

Definition 13 (M-Fidelity)
Let $\mathcal{A}$ and $\mathcal{B}$ be quantum channels. For any $M \in \mathbb{M}_d(\mathbb{C})$, the M-fidelity is

$$f_M(\mathcal{A}, \mathcal{B}) = \frac{\langle \mathcal{A}(M)|\mathcal{B}(M)\rangle}{||M||_2^2}. \tag{2.25}$$

Obviously, monitoring the action of channels on a single matrix won't provide us with a well-rounded picture. It makes more sense to study the M-fidelities over a collection of matrices. Consider any basis of matrices, $\{M_i\}$, orthonormal under the Hilbert-Schmidt inner product. As we have seen, the vectorization of these matrices then becomes an orthonormal basis of vectors. And so, averaging over this collection, we have

$$\begin{aligned}\mathbb{E}(f_{M_i}(\mathcal{A},\mathcal{B})) = \sum_i \frac{1}{d^2} f_{M_i}(\mathcal{A},\mathcal{B}) &= \frac{1}{d^2}\sum_i \frac{\langle \mathcal{A}(M_i)|\mathcal{B}(M_i)\rangle}{||M_i||_2^2} \\ &= \frac{1}{d^2}\sum_i \langle M_i|\boxed{\mathcal{A}}^\dagger\boxed{\mathcal{B}}|M_i\rangle \\ &= \frac{1}{d^2}\mathrm{Tr}(\boxed{\mathcal{A}}^\dagger\boxed{\mathcal{B}}) \\ &= \frac{1}{d^2}\langle \supset\!\!\!\square_{\mathcal{A}}|\supset\!\!\!\square_{\mathcal{B}}\rangle\,.\end{aligned} \tag{2.26}$$

This average — this weighted overlap between Choi matrices — is related to the **entanglement fidelity**, originally studied by Schumacher [29]:

$$F_e(\mathcal{A},|\psi_{EN}\rangle) \equiv \langle\psi_{EN}|I\otimes\mathcal{A}(|\psi_{EN}\rangle\langle\psi_{EN}|)|\psi_{EN}\rangle\,. \tag{2.27}$$

where $|\psi_{EN}\rangle$ is an entangled state.

This quantity was later considered by Nielsen, restricting to the case where $|\psi_{EN}\rangle = |I\rangle/\sqrt{d}$ (the maximally entangled state) [26]:

$$\begin{aligned}F_e(\mathcal{A}) \equiv \frac{1}{d}\langle I|I\otimes\mathcal{A}(\frac{1}{d}|I\rangle\langle I|)|I\rangle &= \frac{1}{d^2}\langle I|\supset\!\!\!\square_{\mathcal{A}}|I\rangle \\ &= \frac{1}{d^2}\mathrm{Tr}(\supset\!\!\!\square_{\mathcal{A}}|I\rangle\langle I|) \\ &= \frac{1}{d^2}\langle \supset\!\!\!\square_{\mathcal{A}}|\supset\!\!\!\square_{\mathcal{I}}\rangle\,.\end{aligned} \tag{2.28}$$

Though he considered only one particular entangled state, Nielsen referred to $F_e(\mathcal{A})$ as the entanglement fidelity of $\mathcal{A}$. For our part, we will be considering (Nielsen) entanglement fidelities of the form $F_e(\mathcal{B}^\dagger\mathcal{A})$.

Definition 14 (Entanglement Fidelity)
Let $\mathcal{A}$ and $\mathcal{B}$ be quantum channels. We define the ***entanglement fidelity between $\mathcal{A}$ and $\mathcal{B}$*** *to be*

$$\Phi(\mathcal{A},\mathcal{B}) \equiv F_e(\mathcal{B}^\dagger\mathcal{A}) = \frac{1}{d^2}\langle \supset\!\!\!\square_{\mathcal{A}}|\supset\!\!\!\square_{\mathcal{B}}\rangle\,. \tag{2.29}$$

To see that the final equality holds in the above equation, note that

$$\begin{aligned} F_e(\mathcal{B}^\dagger\mathcal{A}) &= \frac{1}{d^2}\langle I|I\otimes\mathcal{B}^\dagger\mathcal{A}(|I\rangle\langle I|)|I\rangle \\ &= \frac{1}{d^2}\langle I|I\otimes\mathcal{B}^\dagger(\beth_{\mathcal{A}})|I\rangle \\ &= \frac{1}{d^2}\mathrm{Tr}(I\otimes\mathcal{B}^\dagger(\beth_{\mathcal{A}})\,|I\rangle\langle I|) \\ &= \frac{1}{d^2}\mathrm{Tr}(\sum_k (I\otimes B_k^\dagger)\beth_{\mathcal{A}}(I\otimes B_k)\,|I\rangle\langle I|) \\ &= \frac{1}{d^2}\mathrm{Tr}(\beth_{\mathcal{A}}\sum_k (I\otimes B_k)\,|I\rangle\langle I|\,(I\otimes B_k^\dagger) \\ &= \frac{1}{d^2}\mathrm{Tr}(\beth_{\mathcal{A}}\beth_{\mathcal{B}}) \\ &= \frac{1}{d^2}\langle\beth_{\mathcal{A}}|\beth_{\mathcal{B}}\rangle\,, \end{aligned} \tag{2.30}$$

where, in the final line, we used the fact that Choi matrices of quantum channels are positive semi-definite and thus Hermitian.

As we have expressed, there have been several definitions of entanglement fidelity over the years. In this work, we will use the term "entanglement fidelity" to refer to the two-variable function $\Phi(\mathcal{A},\mathcal{B}) = (1/d^2)\langle\beth_{\mathcal{A}}|\beth_{\mathcal{B}}\rangle$, in definition 14. Since the entanglement fidelity is a real number (equation 2.22), and the Hilbert-Schmidt inner product is conjugated under a swap of its elements, we have that the entanglement fidelity is symmetric in its arguments:

$$\Phi(\mathcal{A},\mathcal{B}) = \Phi(\mathcal{B},\mathcal{A}). \tag{2.31}$$

Also $\Phi(\mathcal{AB},\mathcal{C}) = \mathrm{Tr}(\boxed{\mathcal{AB}}^\dagger\boxed{\mathcal{C}})/d^2 = \mathrm{Tr}(\boxed{\mathcal{B}}^\dagger\boxed{\mathcal{A}}^\dagger\boxed{\mathcal{C}})/d^2$. So

$$\Phi(\mathcal{AB},\mathcal{C}) = \Phi(\mathcal{A},\mathcal{CB}^\dagger) = \Phi(\mathcal{B},\mathcal{A}^\dagger\mathcal{C}) = \Phi(\mathcal{BC}^\dagger,\mathcal{A}^\dagger) = \Phi(\mathcal{ABC}^\dagger,\mathcal{I}). \tag{2.32}$$

It will prove useful to have a term available to describe one argument of the entanglement fidelity, when the other is fixed.

Definition 15 (Target of a Channel)
Let $\mathcal{A}$ and $\mathcal{B}$ be quantum channels. In considering the entanglement fidelity, $\Phi(\mathcal{A},\mathcal{B})$, we refer to $\mathcal{B}$ as the target of $\mathcal{A}$ (and vice-versa).

If, as we've stated, we interpret unitary channels as ideally performed operations, and channels with a larger number of Kraus operators as more noisy processess, then it follows that we would be interested in finding a way to determine how "far away" a given channel is from a unitary one. If we try to apply the gate U, but, due to uncertainty, apply a channel $\mathcal{A}$, then we consider the entanglement fidelity $\Phi(\mathcal{A},\mathcal{U})$. This quantity is related to the **average process fidelity** (a common measure of fidelity for quantum channels), as shown in [26], by:

$$F(\mathcal{A},\mathcal{U}) = \frac{d\Phi(\mathcal{A},\mathcal{U})+1}{d+1}. \tag{2.33}$$

The average process fidelity is a useful measure of a quantum process' performance. It can be estimated experimentally, using well-known techniques like randomized benchmarking [11, 8, 19, 22, 23, 24] and direct fidelity estimation [15, 7]. Since it is completely captured through the entanglement fidelity, our investigation of $\Phi(\mathcal{A},\mathcal{U})$ is not without its practical benefits.

2.3.1 Unitarity

A second quantum process measure that we are interested in, is the entanglement fidelity between a channel and itself: $\Phi(\mathcal{A},\mathcal{A}) = \Phi(\mathcal{A}^\dagger\mathcal{A},\mathcal{I})$. In the case of a unitary channel, the composition $\mathcal{U}^\dagger\mathcal{U} = \mathcal{I} = \mathcal{U}\mathcal{U}^\dagger$. However, for more general channels, this does not hold. Thus, determining how close the channel $\mathcal{A}^\dagger\mathcal{A}$ is to the identity tells us something about how reversible our operation is. It turns out that $\Phi(\mathcal{A},\mathcal{A})$ is related to the concept of the **unitarity** of the channel, $\mathcal{A}$, defined in [31]:

$$u(\mathcal{A}) = \frac{d^2\Phi(\mathcal{A},\mathcal{A})-1}{d^2-1}, \tag{2.34}$$

which, (as the name implies) is a measure of how close a given channel is to a unitary one. Since $\Phi(\mathcal{A},\mathcal{A})$ only depends on $\mathcal{A}$, for ease of notation we will usually write it as the single-variable function, $\Upsilon(\mathcal{A}) \equiv \Phi(\mathcal{A},\mathcal{A})$.

Note: We will sometimes refer to $\Upsilon(\mathcal{A})$ as the unitarity of $\mathcal{A}$ and $\Phi(\mathcal{A},\mathcal{U})$ as the fidelity, when no confusion will arise.

2.4 Depolarizing Channels

Before moving on to the development of the leading Kraus approximation, we would like to look at a particular class of quantum channels, known as **depolarizing channels**:

$$\mathcal{D}_p(M) = pM + (1-p)\mathrm{Tr}(M)I/d. \tag{2.35}$$

In the next chapter, when we explore the concept of **decoherent behaviour**, at length, depolarizing channels should always be in the back of our minds. In some sense, the problem of defining decoherent behaviour is based on the idea of generalizing depolarizing channels.

The Choi matrix of such a channel is

$$\begin{aligned}
\beth_{\mathcal{D}} &= \sum_{ij} E_{ij} \otimes (pE_{ij} + (1-p)\delta_{ij}I/d) \\
&= p\left|I\right\rangle\left\langle I\right| + (1-p)\sum_{i} E_{ii} \otimes I/d \\
&= p\left|I\right\rangle\left\langle I\right| + \frac{(1-p)}{d} I \otimes I.
\end{aligned} \tag{2.36}$$

Taking the inner product of Choi matrices, we have

$$\begin{aligned}
\left\langle \beth_{\mathcal{D}_p} \middle| \beth_{\mathcal{A}} \right\rangle &= \mathrm{Tr}(\beth_{\mathcal{D}_p}\beth_{\mathcal{A}}) \\
&= \mathrm{Tr}(p\left|I\right\rangle\left\langle I\right|\beth_{\mathcal{A}} + \frac{(1-p)}{d}\beth_{\mathcal{A}}) \\
&= p\mathrm{Tr}(\sum_{ijkl} E_{ij}E_{kl} \otimes E_{ij}\mathcal{A}(E_{kl})) + (1-p). \\
&= p\sum_{ij} \mathrm{Tr}(E_{ij}\mathcal{A}(E_{ji})) + (1-p) \\
&= p\sum_{ijk} \left\langle j|A_k|j\right\rangle \left\langle i|A_k^\dagger|i\right\rangle + (1-p) \\
&= p\sum_{k} \left|\mathrm{Tr}(A_k)\right|^2 + (1-p).
\end{aligned} \tag{2.37}$$

Dividing by d^2, we arrive at an expression for the entanglement fidelity: $\Phi(\mathcal{D},\mathcal{A}) = p\Phi(\mathcal{A},\mathcal{I}) + (1-p)/d^2$. This is a neat result. The second term corresponds to a loss of information: $(1-p)$ is the probability that *any* state is mapped to I/d. It is interesting that there actually is some amount of information that is preserved in this term; $(1-p)/d^2$

decreases quickly as the size of the system is increased (as more information is lost), but it is non-zero. The information that is preserved comes from the trace-preservation of $\mathcal{D}_p$ and $\mathcal{A}$. This term is also independent of the target channel, $\mathcal{A}$; trace preservation is the best we can do to preserve/recover information from this term. On the other hand, the first term does allow us recovery, and in fact we will recover the most information when our target channel is the identity. In chapter 3, we will see that the largest fidelity for any (wide sense equable) decoherent channel is approximately obtained with an identity target channel.

We arrive at a fidelity of $\Phi(\mathcal{D},\mathcal{I}) = p + (1-p)/d^2$, and unitarity, $\Upsilon(\mathcal{D}) = p^2 + (1-p)(1+p)/d^2$. We see that $\Phi(\mathcal{D},\mathcal{I}) \approx p \approx \sqrt{\Upsilon(\mathcal{D})}$. A quadratic relationship appears between the unitarity and fidelity. Interestingly, in a way, this relationship transcends depolarizing channels. In chapter 3, we will see that any decoherent channel satisfying a few conditions will have such a quadratic relationship between its fidelity and unitarity. As a final result, consider a sequence of depolarizing channels, with depolarizing constants, p_i: $\mathcal{D}_i(M) = p_i M + (1-p_i)\mathrm{Tr}(M)I/d$.

$$\begin{aligned}
\mathcal{D}_2\mathcal{D}_1(M) &= \mathcal{D}_2(p_1 M + (1-p_1)\mathrm{Tr}(M)I/d) \\
&= p_2 p_1 M + (1-p_2)p_1\mathrm{Tr}(M)I/d + p_2(1-p_1)\mathrm{Tr}(M)I/d \\
&\quad + (1-p_1)(1-p_2)\mathrm{Tr}(M)I/d \\
&= p_2 p_1 M + (p_1 - p_1p_2 + p_2 - p_1p_2 + 1 - p_1 - p_2 + p_1p_2)\mathrm{Tr}(M)I/d \\
&= p_2 p_1 M + (1 - p_2p_1)\mathrm{Tr}(M)I/d.
\end{aligned} \tag{2.38}$$

Turtling down, we find

$$\mathcal{D}_m \ldots \mathcal{D}_1(M) = \prod_i p_i M + (1 - \prod_i p_i)\mathrm{Tr}(M)I/d. \tag{2.39}$$

Thus the product of a set of depolarization channels is a depolarization channel whose depolarization constant is the product of those in the composition. And so we find a level of structure in the set of depolarizing channels — closure under composition. In the final section of chapter 3, we will extend this result to the closure of the set of **decoherence-limited** channels. Furthermore, we find

$$\Phi(\mathcal{D}_m \ldots \mathcal{D}_1, \mathcal{I}) = \prod_i p_i + \frac{1}{d^2}(1 - \prod_i p_i) \approx \prod_i p_i \approx \prod_i \Phi(\mathcal{D}_i, \mathcal{I}). \tag{2.40}$$

We will see something very similar in theorem 3.6.2.

Chapter 3

The Leading Kraus Approximation

We are now prepared to develop the leading Kraus approximation. Recall that given any channel, $\mathcal{A}$, the eigenvectors of the Choi matrix, $\beth_{\mathcal{A}}$, provide an orthonormal set of Kraus operators. These *canonical Kraus operators* will be the only Kraus operators we shall use, moving forward. And so from here on in, "Kraus operator" should be read "canonical Kraus operator", unless otherwise specified.

Definition 16 (Leading Kraus Approximation)
Given a channel, $\mathcal{A}$, order the Kraus operators using the Frobenius norm:

$$\|A_1\|_F \geq \|A_2\|_F \geq \dots \|A_{d^2}\|_F \geq 0. \tag{3.1}$$

*We call A_1 the **leading Kraus operator** of $\mathcal{A}$.*
*The **leading Kraus approximation** (LKA) of $\mathcal{A}$ is:*

$$\mathcal{A}^{\circ}(\rho) \equiv A_1 \rho A_1^{\dagger}\ , \quad \forall \rho \in \mathbb{M}_d(\mathbb{C}). \tag{3.2}$$

Since we will be using the canonical Kraus operators, we can alternatively write equation 3.1 in terms of the eigenvalues of the Choi matrix, $\beth_{\mathcal{A}}$:

$$\lambda_{A,1} \geq \lambda_{A,2} \geq \dots \geq \lambda_{A,d^2}. \tag{3.3}$$

Note: The LKA is a completely positive channel, but it is trace preserving iff $\mathcal{A}$ is a unitary channel. Thus performing the LKA will not preserve the physicality of most channels. However, as we shall see, the leading Kraus operator closely describes the average behaviour of the full channel.

It is possible to construct channels in which the leading Kraus operator is not unique. In what follows, we would like to be able to speak of *the* leading Kraus operator for each channel. There is a simple condition which we can enforce, which will eliminate the possibility of degeneracy in our leading Kraus operators. To arrive at the required condition, we consider the entanglement fidelity, $\Phi(\mathcal{A}, \mathcal{B})$. Expanding the Choi matrices into their spectral decompositions, $\beth_{\mathcal{A}} = \sum_j \lambda_{A,j} |\hat{a}_j\rangle \langle \hat{a}_j|$, we have:

$$\begin{aligned}
\Phi(\mathcal{A}, \mathcal{B}) &= \frac{1}{d^2} \mathrm{Tr}(\beth_A \beth_B) \\
&= \frac{1}{d^2} \mathrm{Tr}(\sum_{jk} \lambda_{A,j} \lambda_{B,k} \left|\hat{a}_j\right\rangle \left\langle \hat{a}_j \middle| \hat{b}_k \right\rangle \left\langle \hat{b}_k\right|) \\
&= \frac{1}{d^2} \sum_{jk} \lambda_{A,j} \lambda_{B,k} \left| \left\langle \hat{a}_j \middle| \hat{b}_k \right\rangle \right|^2.
\end{aligned} \tag{3.4}$$

We can then upper bound the entanglement fidelity by upper bounding each eigenvalue of $\beth_{\mathcal{A}}$ by $\lambda_{A,1}$:

$$\begin{aligned}
\Phi(\mathcal{A}, \mathcal{B}) &= \frac{1}{d^2} \sum_{jk} \lambda_{A,j} \lambda_{B,k} \left| \left\langle \hat{a}_j \middle| \hat{b}_k \right\rangle \right|^2 \\
&\leq \frac{\lambda_{A,1}}{d^2} \sum_{jk} \lambda_{B,k} \left| \left\langle \hat{a}_j \middle| \hat{b}_k \right\rangle \right|^2 \\
&= \frac{\lambda_{A,1}}{d^2} \sum_{k} \lambda_{B,k} \left\langle \hat{b}_k \middle| \hat{b}_k \right\rangle \\
&= \lambda_{A,1}/d.
\end{aligned} \tag{3.5}$$

In the third line we used the fact that the eigenvectors of the Choi matrix $\beth_{\mathcal{A}}$ form an orthonormal basis, and thus their outer-product sums to the identity.

Since the entanglement fidelity is symmetric, we also have $\Phi(\mathcal{A}, \mathcal{B}) \leq \lambda_{B,1}/d$. Thus by enforcing $\Phi(\mathcal{A}, \mathcal{B}) > 1/2$, we have $\lambda_{A,1}, \lambda_{B,1} > d/2$. Knowing that the eigenvalues of Choi matrices are non-negative and sum to d for quantum channels, we will have a unique largest eigenvalue of the Choi matrices of both $\mathcal{A}$ and $\mathcal{B}$ (and thus unique leading Kraus operators) if $\Phi(\mathcal{A}, \mathcal{B}) > 1/2$.

So given a channel, $\mathcal{A}$, all we need is for $\Phi(\mathcal{A}, \mathcal{B}) > 1/2$ for *some* $\mathcal{B}$. Later on in this chapter, we will see that it will be useful for us to require the entanglement fidelity to surpass $1/2$ for two different channels — $\mathcal{A}$ and $\mathcal{U}$ — as this will allow us to draw conclusions about the unitarity and fidelity, respectively. Before proceeding to those results, we present a lemma which will be very useful in the proof of theorem 3.1.1.

Lemma 3.0.1
Let $\{p_i\}_{i=1}^n$ and $\{q_i\}_{i=1}^n$ be probability distributions:

$$\sum_i p_i = \sum_i q_i = 1 \ ; \ p_i, q_i \geq 0 \ , \ \forall i. \tag{3.6}$$

If $\sum_i p_i q_i > 1/2$ then $\{p_i\}$ and $\{q_i\}$ have unique largest elements, $p_{\max}$ and $q_{\max}$, which occur at the same index.

Proof. By symmetry, we can just frame everything in terms of $\{p_i\}$. If $p_{\max} \leq 1/2$, we have

$$\sum_j p_j q_j \leq p_{\max} \sum_j q_j = p_{\max} \leq 1/2. \tag{3.7}$$

Thus for $\sum_i p_i q_i > 1/2$ to occur, $p_{\max} > 1/2$ and so there is a unique largest element of $\{p_i\}$. Wlog, say that $p_1 = p_{\max}$.

Let $p_{\max} = 1/2 + \epsilon_p$ with $0 < \epsilon_p \leq 1/2$. Assume $q_{\max} \neq q_1$. Then we must have $0 \leq q_1 \leq 1/2$. So we can write $q_1 = 1/2 - \epsilon_q$ with $0 \leq \epsilon_q \leq 1/2$. Then

$$\begin{aligned}
\sum_j p_j q_j &= p_{\max} q_1 + \sum_{j \neq 1} p_j q_j \\
&\leq p_{\max} q_1 + (\sum_{j \neq 1} p_j)(\sum_{j \neq 1} q_j) \\
&= p_{\max} q_1 + (1 - p_{\max})(1 - q_1) \\
&= 2 p_{\max} q_1 + 1 - p_{\max} - q_1 \\
&= 2(1/4 + (1/2)\epsilon_p - (1/2)\epsilon_q - \epsilon_p \epsilon_q) + 1 - 1/2 - \epsilon_p - 1/2 + \epsilon_q \\
&= 1/2 - 2\epsilon_p \epsilon_q \\
&\leq 1/2.
\end{aligned} \tag{3.8}$$

So by contradiction, the maximum terms of $\{p_i\}$ and $\{q_i\}$ must occur on the same index. ■

Note: The form for the entanglement fidelity is $\Phi(\mathcal{A}, \mathcal{B}) = (1/d^2) \sum_{jk} \lambda_{A,j} \lambda_{B,k} \left| \left\langle \hat{a}_j \middle| \hat{b}_k \right\rangle \right|^2$. We also have

$$\sum_j \frac{\lambda_{A,j}}{d} = 1. \tag{3.9a}$$

$$\begin{aligned}\frac{1}{d}\sum_j\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_j\middle|\hat{b}_k\right\rangle\right|^2 &= \frac{1}{d}\sum_k \lambda_{B,k}\left\langle \hat{b}_k\middle|\sum_j \left|\hat{a}_j\right\rangle\left\langle \hat{a}_j\middle|\hat{b}_k\right\rangle\right. \\ &= \frac{1}{d}\sum_k \lambda_{B,k}\left\langle \hat{b}_k\middle|\hat{b}_k\right\rangle \\ &= 1.\end{aligned} \tag{3.9b}$$

So the entanglement fidelity is the product of two probability distributions: $\{\lambda_{A,i}/d\}$ and $\{\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_j\middle|\hat{b}_k\right\rangle\right|^2/d\}$.

Note: We use the term "probability distribution", quite liberally here, to refer to a set of non-negative values that sum to one, even if those values do not necessarily correspond to probabilities of particular events; much like the use of the term "probability" in "probability vector", which is just a vector of non-negative entries which sum to one. Of course, as we have already established, the distribution $\{\lambda_{A,i}/d\}$ can be thought of as the probabilities of the corresponding canonical Kraus operators being applied in the action of the channel $\mathcal{A}$.

By the above lemma, as long as the entanglement fidelity is greater than $1/2$, the largest terms of the two distributions will have matching indices. And so $\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_1\middle|\hat{b}_k\right\rangle\right|^2$ will be the largest inner product term. This leads to

$$\begin{aligned}\Phi(\mathcal{A},\mathcal{B}) &= \frac{1}{d^2}\sum_{jk}\lambda_{A,j}\lambda_{B,k}\left|\left\langle \hat{a}_j\middle|\hat{b}_k\right\rangle\right|^2 \\ &\leq \frac{1}{d^2}\sum_{jk}\lambda_{A,j}\lambda_{B,k}\left|\left\langle \hat{a}_1\middle|\hat{b}_k\right\rangle\right|^2 \\ &= \frac{1}{d}\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_1\middle|\hat{b}_k\right\rangle\right|^2,\end{aligned} \tag{3.10}$$

which will be useful in proving the next theorem.

3.1 Preliminary Results

Theorem 3.1.1
Let $\mathcal{A}$ and $\mathcal{B}$ be quantum channels with $\Phi(\mathcal{A},\mathcal{B}) > 1/2$. Then

$$0 \leq \Phi(\mathcal{A},\mathcal{B}) - \frac{\lambda_{A,1}}{d^2}\langle \hat{a}_1|\beth_{\mathcal{B}}|\hat{a}_1\rangle \leq (1-\Upsilon(\mathcal{A}))(1-\Phi(\mathcal{A},\mathcal{B})). \tag{3.11}$$

Or in terms of the Kraus operators,

$$0 \leq \Phi(\mathcal{A},\mathcal{B}) - \sum_k \left|\frac{\mathrm{Tr}(A_1^\dagger B_k)}{d}\right|^2 \leq (1-\Upsilon(\mathcal{A}))(1-\Phi(\mathcal{A},\mathcal{B})). \tag{3.12}$$

Proof.

$$\begin{aligned}
\Phi(\mathcal{A},\mathcal{B}) - \frac{\lambda_{A,1}}{d^2}\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_1\middle|\hat{b}_k\right\rangle\right|^2 &= \frac{1}{d^2}\sum_{j\neq 1}\sum_k \lambda_{A,j}\lambda_{B,k}\left|\left\langle \hat{a}_j\middle|\hat{b}_k\right\rangle\right|^2 \\
&\leq \frac{1}{d^2}\Big(\sum_{j\neq 1}\lambda_{A,j}\Big)\Big(\sum_{j\neq 1}\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_j\middle|\hat{b}_k\right\rangle\right|^2\Big) \\
&= \Big(1-\frac{\lambda_{A,1}}{d}\Big)\Big(1-\frac{1}{d}\sum_k \lambda_{B,k}\left|\left\langle \hat{a}_1\middle|\hat{b}_k\right\rangle\right|^2\Big) \\
&\leq (1-\Upsilon(\mathcal{A}))(1-\Phi(\mathcal{A},\mathcal{B})).
\end{aligned} \tag{3.13}$$

Where, in the last line, we used equations 3.5 and 3.10. ■

Two results immediately follow, providing physical intuition for the above theorem, restricting to the cases of the unitarity and average process fidelity.

Corollary 3.1.1
Let $\mathcal{A}$ be a quantum channel with $\Upsilon(\mathcal{A}) > 1/2$. Then

$$0 \leq \Upsilon(\mathcal{A}) - \left(\frac{\lambda_{A,1}}{d}\right)^2 \leq (1-\Upsilon(\mathcal{A}))^2 \tag{3.14}$$

Proof. Using theorem 3.1.1 and the fact that

$$\frac{\lambda_{A,1}}{d^2}\langle \hat{a}_1|\beth_{\mathcal{A}}|\hat{a}_1\rangle = \frac{\lambda_{A,1}}{d^2}\sum_k \lambda_{A,k}|\langle \hat{a}_1|\hat{a}_k\rangle|^2 = (\frac{\lambda_{A,1}}{d})^2 \tag{3.15}$$

■

Corollary 3.1.2
Let $\mathcal{A}$ and $\mathcal{U}$ be quantum channels with $\Phi(\mathcal{A},\mathcal{U}) > 1/2$. Then

$$0 \leq \Phi(\mathcal{A},\mathcal{U}) - \frac{\lambda_{A,1}}{d}|\langle \hat{a}_1|\hat{u}\rangle|^2 \leq (1-\Upsilon(\mathcal{A}))(1-\Phi(\mathcal{A},\mathcal{U})). \tag{3.16}$$

Proof. Recall that unitary channels only have a single Kraus operator. So $\beth_{\mathcal{U}}$ must have only one non-zero eigenvalue, which is equal to d. Thus,

$$\frac{\lambda_{A,1}}{d^2}\langle \hat{a}_1|\beth_{\mathcal{U}}|\hat{a}_1\rangle = \frac{\lambda_{A,1}}{d^2}d|\langle \hat{a}_1|\hat{u}\rangle|^2 = \frac{\lambda_{A,1}}{d}|\langle \hat{a}_1|\hat{u}\rangle|^2. \tag{3.17}$$

■

Note: As we continue through this chapter, we will find many upper bounds of the form $(1-X)^2$ showing up in our results. In the literature, fidelities over 0.9 are typically discussed [20, 14, 1]. As such, these upper bounds will be quite small, and the leading Kraus operator will provide a good approximation for the quantities of interest.

We can see that with modest assumptions on our channel, the leading Kraus operator will capture the majority of the unitarity and fidelity. We view the leading Kraus operator as accounting for most of the average behaviour of our (non-catastrophic) channels, while the other operators provide corrections. With our pair of corollaries as inspiration, we define the class of channels that we will continue to work with for the remainder of this chapter.

Definition 17 (Non-Catastrophic Channels)
*A quantum channel, $\mathcal{A}$, is called **non-catastrophic** if*

$$\Upsilon(\mathcal{A}) > 1/2 \tag{3.18a}$$

and

$$\Phi(\mathcal{A},\mathcal{U}) > 1/2, \tag{3.18b}$$

*for some $\mathcal{U}$, called the **target** of $\mathcal{A}$.*

We next seek the behaviour of the leading Kraus approximation under compositions of channels. Given a block of composed channels, $\mathcal{A}_1\mathcal{A}_2\ldots\mathcal{A}_n$, if we take the leading Kraus operator of each channel and compose them, $\mathcal{A}_1^\circ\mathcal{A}_2^\circ\ldots\mathcal{A}_n^\circ$, how much of the unitarity and fidelity will be captured? What additional assumptions will we need to make about our channels? This is the topic of the next section.

3.2 Composition Results

Let us first introduce some notation to simplify the upcoming results:

Definition 18 (Composition Notation)
Let $\{\mathcal{A}_i\}_{i=1}^m$ *be a set of matrices. We define:*

$$\mathcal{A}_{m:1} \equiv \mathcal{A}_m \mathcal{A}_{m-1} \dots \mathcal{A}_1 \tag{3.19}$$

and

$$\mathcal{A}^\circ_{m:1} \equiv \mathcal{A}^\circ_m \mathcal{A}^\circ_{m-1} \dots \mathcal{A}^\circ_1. \tag{3.20}$$

That is, $\mathcal{A}^\circ_{m:1}$ *is the composition of the leading Kraus operators of* $\mathcal{A}_1, \dots, \mathcal{A}_m$*, NOT the leading Kraus operator of the composition.*

If we are able to obtain results on the behaviour of leading Kraus operators, with respect to compositions, we will be provided some understanding of the evolution of quantum channels; viewing composition as discrete time evolution. Using theorem 3.1.1, we immediately have results involving the leading Kraus operator of a composition. Unfortunately, it is not immediately obvious, given information about the composed channels, what the leading Kraus operator of the composition is. When a channel is composed with a unitary channel, $\mathcal{UA}$, the canonical Kraus operators will just be $\{UA_i\}$, where $\{A_i\}$ are the canonical Kraus operators of $\mathcal{A}$. This can be seen through:

$$\begin{aligned} \beth_{\mathcal{UA}} &= \sum_{ij} E_{ijk} \otimes U A_k E_{ij} A_k^\dagger U^\dagger \\ &= (I \otimes U) \beth_{\mathcal{A}} (I \otimes U)^\dagger. \end{aligned} \tag{3.21}$$

Since $I \otimes U$ is unitary, $\beth_{\mathcal{UA}}$ has the same eigenvalues as $\beth_{\mathcal{A}}$, with eigenvectors $(I \otimes U) |\hat{a}_i\rangle = |U\hat{a}_i\rangle$. Thus the canonical Kraus operators will be the unvectorized $\sqrt{\lambda_{A,i}} |U\hat{a}_i\rangle$: i.e. $\{UA_i\}$, with LKO, UA_1.

Similarly,

$$\begin{aligned} \beth_{\mathcal{AU}} &= \sum_k (I \otimes A_k) \beth_{\mathcal{U}} (I \otimes A_k)^\dagger \\ &= \sum_k (I \otimes A_k) d \, |\hat{u}\rangle \langle \hat{u}| \, (I \otimes A_k)^\dagger \\ &= \sum_k d \, |A_k \hat{u}\rangle \langle A_k \hat{u}| \, , \end{aligned} \tag{3.22}$$

where $\hat{u} = U/\sqrt{d}$. Noting that $\langle A_i\hat{u}|A_j\hat{u}\rangle = \mathrm{Tr}(UA_i^\dagger A_j U^\dagger)/d = (\lambda_{A,i}/d)\delta_{ij}$, we see that the eigenvalues of $\beth_{\mathcal{AU}}$ are $\lambda_{A,i}$, with normalized eigenvectors

$$\frac{|A_i\hat{u}\rangle}{\sqrt{\langle A_i\hat{u}|A_i\hat{u}\rangle}} = \frac{\sqrt{d}}{\sqrt{\lambda_{A,i}}}\,|A_i\hat{u}\rangle = |\hat{a}_i U\rangle\,. \tag{3.23}$$

So the canonical Kraus operators are $\{A_iU\}$, with LKO, A_1U.

In general, however, the leading Kraus operators of compositions of channels cannot be easily found from the LKOs of the composed terms: $(\mathcal{A}_{m:1})^\circ \neq \mathcal{A}^\circ_{m:1}$. With a bit of work, however, we can arrive at a pair of useful theorems.

Note: The proofs for the remaining results in this chapter are much more involved than the ones we have thus far encountered. We shall state the remaining results without proof (except for a quick result in theorem 3.2.1) and direct the reader to [3] for their derivations, as we do not believe that we can provide any additional insight by presenting the proofs here.

Theorem 3.2.1 (LKA - Composition Unitarity)
Consider m non-catastrophic channels, $\mathcal{A}_i$, and suppose that $\mathcal{A}_{m:1}$ is also non-catastrophic. Then

$$0 \leq \Upsilon(\mathcal{A}_{m:1}) - \Upsilon(\mathcal{A}^\circ_{m:1}) \leq (1 - \Upsilon(\mathcal{A}_{m:1}))^2 \leq (1 - \Upsilon(\mathcal{A}^\circ_{m:1}))^2. \tag{3.24}$$

Proof. The first and final inequality arise from the fact that each term in the sum $\Upsilon(\mathcal{A}_{m:1}) = \sum_{i_1\ldots i_m} |\mathrm{Tr}(A_{m,i_m}\ldots A_{1,i_1})|^2/d^2$ is non-negative and $\Upsilon(\mathcal{A}^\circ_{m:1})$ contains only one of these terms.

■

Theorem 3.2.2 (LKA - Composition Fidelity)
Consider m non-catastrophic channels $\mathcal{A}_i$ with respective unitary targets $\mathcal{U}_i$ and suppose that the composition $\mathcal{A}_{m:1}$ is also non-catastrophic. Then,

$$0 \leq \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}^{\circ}_{m:1}, \mathcal{U}_{m:1}) < \frac{1}{2}\left(\sum_{i=1}^{m}(1 - \Upsilon(\mathcal{A}^{\circ}_i))\right)^2 + (1 - \Phi(\mathcal{A}^{\circ}_{m:1}, \mathcal{U}_{m:1}))\sum_{i=1}^{m}(1 - \Upsilon(\mathcal{A}^{\circ}_i)) \ , \tag{3.25}$$

and

$$0 \leq \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) - \Phi(\mathcal{A}^{\circ}_{m:1}, \mathcal{U}_{m:1}) < \frac{1}{2}\left(\sum_{i=1}^{m}(1 - \Upsilon(\mathcal{A}_i))\right)^2 + (1 - \Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}))\sum_{i=1}^{m}(1 - \Upsilon(\mathcal{A}_i)) + \text{ H.O.} \tag{3.26}$$

Note: We use "H.O." to denote higher order terms.

For sufficiently well-behaved (non-catastrophic) channels, the utility of the leading Kraus operator persists beyond the channel itself. If we view composition as a discrete time evolution,

$$\mathcal{A}(t_n) = \mathcal{A}_{n:1}, \tag{3.27}$$

where each channel inches the state further in time, then the leading Kraus operators are the principle actors, and we view their action like tracing the path of the center of mass.

We would like to press forward from discrete time evolution to the continuous. However, before we do, we would like to leverage the leading Kraus approximation to derive a decomposition for non-catastrophic channels. This decomposition will give rise to a notion of *decoherent channels*, the generalization of depolarizing channels we mentioned in chapter 2. We shall find that our definition of decoherent channels fits in nicely with well-known ideas in the study of the **Lindblad equation** [21, 16], an equation describing the evolution of quantum states in time.

3.3 Decoherent Channels

Definition 19 (Decoherent Channels)
A non-catastrophic channel is called ***decoherent*** *if its Leading Kraus operator is positive semi-definite.*

Definition 20 (Polar Decomposition of Channels)
Let $\mathcal{A}$ be a non-catastrophic channel. We can decompose $\mathcal{A}$ as

$$\mathcal{A} = \mathcal{U}\mathcal{D}. \tag{3.28a}$$

$$\mathcal{A} = \mathcal{D}'\mathcal{U}. \tag{3.28b}$$

Where the Kraus operator of $\mathcal{U}$ is the unitary part of the polar decomposition of the LKO: $A_1 = UP$. The Kraus operators of $\mathcal{D}$ are $\{U^\dagger A_i\}$ and the Kraus operators of $\mathcal{D}'$ are $\{A_i U^\dagger\}$.

$$\mathcal{D}^\circ(\rho) = |A_1|\rho|A_1|^\dagger. \tag{3.29a}$$

$$\mathcal{D}'^\circ(\rho) = U|A_1|U^\dagger\rho U|A_1|^\dagger U^\dagger. \tag{3.29b}$$

Note: $\mathcal{D}$ and $\mathcal{D}'$ are completely positive, simply through having Kraus operators. Trace preservation holds by:

$$\sum_i (U^\dagger A_i)^\dagger U^\dagger A_i = \sum_i A_i^\dagger A_i = I = UIU^\dagger = \sum_i (A_i U^\dagger)^\dagger A_i U^\dagger. \tag{3.30}$$

Let us consider the types of errors we could expect in a realistic implementation of a quantum algorithm. As a simple sorting, we can divide our channels based on whether or not they can be repaired through unitary action. Any unitary error can be removed simply by applying the inverse: $\mathcal{U}^\dagger\mathcal{U} = \mathcal{I}$. Theorem 3.5.2 and theorem 3.8.1 (which we will see later) tell us that this is, in some sense, all there is to be said about unitary reparation — fidelities are maximized by removing the unitary components from polar decompositions.

Unitary errors are also called **coherent errors**. These appear when we find ourselves having mishandled our quantum system (leading to the application of an incorrect gate), but we haven't introduced any additional uncertainty; if we find our mistake, we can correct it. **Decoherent errors** on the other hand, correspond to irreversible losses of information. This was the type of behaviour we saw for depolarizing channels, in section 2.4: $\Phi(\mathcal{D}_p, \mathcal{A}) = p\Phi(\mathcal{A}, \mathcal{I}) + (1-p)/d^2$. There is no unitary we can apply to correct the fidelity of our channel. In fact the greatest fidelity we can achieve is by simply letting $\mathcal{A} = \mathcal{I}$. We shall see other behaviours of decoherent channels, analogous to those of

depolarizing channels, in theorem 3.5.2, in which we find $\Phi(\mathcal{D}, \mathcal{I}) \approx \sqrt{\Upsilon(\mathcal{D})}$, and theorem 3.6.2, which gives $\Phi(\mathcal{D}_{m:1}, I) \approx \prod_i \Phi(\mathcal{D}_i, I)$, for depolarizing channels, $\{\mathcal{D}_i\}$. Our polar decomposition, then, splits the behaviour of a (non-catastrophic) quantum channel into coherent and decoherent contributions, both intuitive.

3.4 The Lindblad Equation

We let our state depend on time: $\rho(t)$. An infinitesimal progression in time can be seen as arising from a quantum channel:

$$\rho(t + dt) = \sum_k A_k(t, dt)\rho(t)A_k^\dagger(t, dt). \tag{3.31}$$

We won't prove it here, but any invertible matrix can be written as an exponential of some other matrix, where

$$e^A \equiv \sum_n \frac{1}{n!}(A)^n. \tag{3.32}$$

For the proof, see for instance [17]. Any small perturbation will transform a singular matrix into an invertible one, so we shall assume that our Kraus operators will remain non-singular, and we write $A_1(t) = \exp(B(t))$. Since dt is infinitesimal, we write

$$\begin{aligned} A_1(t, dt) &= \exp(-iH(t)dt - P(t)dt) \\ &= I - iH(t)dt - P(t)dt + O(dt^2), \end{aligned} \tag{3.33}$$

where $H(t)$ is Hermitian for all t, and we define $P(t)dt \equiv -B(t)dt - iH(t)dt$.

Since the channel, $\mathcal{A}(t)$, will be trace preserving at all times, we have

$$\begin{aligned} I &= \sum_k A_k^\dagger(t, dt)A_k(t, dt) \\ &= A_1(t, dt)^\dagger A_1(t, dt) + \sum_{k\neq 1} A_k^\dagger(t, dt)A_k(t, dt) \\ &= I - iH(t)dt + iH(t)dt - 2P(t)dt + \sum_{k\neq 1} A_k^\dagger(t, dt)A_k(t, dt) + O(dt^2). \end{aligned} \tag{3.34}$$

So

$$P(t)dt = \frac{1}{2}\sum_{k\neq 1} A_k^\dagger(t, dt)A_k(t, dt) + O(dt^2). \tag{3.35}$$

On the left hand side we have something scaling as dt. The higher order terms on the right hand side are order dt^2, so the linear, dt, terms must be in the sum. So this implies that the non-leading Kraus operators must all be order $\sqrt{dt}$. Furthermore, we see $P(t)$ is positive-semidefinite at all times (justifying the notation).

$$\begin{aligned}
\rho(t+dt)-\rho(t) &= \sum_k A_k(t,dt)\rho(t)A_k^\dagger(t,dt)-\rho(t) \\
&= (I-iH(t)dt-P(t)dt+O(dt^2))\rho(t)(I+iH(t)dt-P(t)dt+O(dt^2)) \\
&\quad +\sum_{k\neq 1} A_k(t,dt)\rho(t)A_k^\dagger(t,dt)-\rho(t) \\
&= \rho(t)-iH(t)\rho(t)dt+i\rho(t)H(t)dt-P(t)\rho(t)dt-\rho(t)P(t)dt \\
&+\sum_{k\neq 1} A_k(t,dt)\rho(t)A_k^\dagger(t,dt)-\rho(t)+O(dt^2) \\
&= -i[H(t),\rho(t)]dt-\{P(t),\rho(t)\}dt+\sum_{k\neq 1} A_k(t,dt)\rho(t)A_k^\dagger(t,dt)+O(dt^2)
\end{aligned} \tag{3.36}$$

where $[A,B]=AB-BA$ and $\{A,B\}=AB+BA$ are the commutator and anticommutator, respectively.

"Dividing" by dt and sending $dt\to 0$, gives us

$$\frac{d\rho}{dt}=-i[H(t),\rho(t)]-\frac{1}{2}\left\{\sum_{k\neq 1}L_k^\dagger(t)L_k(t),\rho(t)\right\}+\sum_{k\neq 1}L_k(t)\rho(t)L_k^\dagger(t), \tag{3.37}$$

where

$$L_k(t)\equiv\lim_{dt\to 0}\frac{A_k(t,dt)}{\sqrt{dt}} \tag{3.38}$$

are the **Lindblad operators** and

$$P(t)=\frac{1}{2}\sum_{k\neq 1}L_k^\dagger(t)L_k(t). \tag{3.39}$$

Equation 3.37 is the **Lindblad equation**.

Using the orthonormality of the canonical Kraus operators and equation 3.33, we have

$$\begin{aligned}
0 &= \mathrm{Tr}(A_1^\dagger(t,dt)A_{k\neq 1}(t,dt)) \\
&= \mathrm{Tr}(A_{k\neq 1}(t,dt))+dt\big[i\mathrm{Tr}(H(t)A_{k\neq 1}(t,dt))-\mathrm{Tr}(P(t)A_{k\neq 1}(t,dt))\big]+O(dt^2\sqrt{dt}).
\end{aligned} \tag{3.40}$$

Dividing through by $\sqrt{dt}$ and taking the limit as $t \to 0$ (noting that the trace is a continuous function) we arrive at

$$\mathrm{Tr}(L_k(t)) = 0, \ \forall k \neq 1. \tag{3.41}$$

Note: It is through the canonical Kraus operators that we have been able to force the Lindblad operators to have zero trace at all times. If we had chosen a different set of Kraus operators, equation 3.41 would not hold.

Taking a closer look at the Lindblad equation, we examine each of the three terms:

$$\frac{d\rho}{dt} = \underbrace{-i[H(t), \rho(t)]}_{1} \underbrace{- \frac{1}{2}\left\{\sum_{k\neq 1} L_k^\dagger(t) L_k(t), \rho(t)\right\}}_{2} + \underbrace{\sum_{k\neq 1} L_k(t)\rho(t)L_k^\dagger(t)}_{3}. \tag{3.42}$$

The first term in this equation corresponds to coherent behaviour. This might be easier to see once we note that $\exp(iH)$ is unitary, for any Hermitian matrix, H. In the absence of $P(t)$, then, equation 3.33 tells us that A_1 will be unitary. The second and third terms correspond to decoherent behaviour, and are often referred to as the relaxation or decoherent part of the Lindbladian [18, 12].

Looking at equation 3.33 and taking the conjugate transpose, we have

$$A_1^\dagger(t, dt) = I + iH(t)dt - P(t)dt + O(dt^2). \tag{3.43}$$

So for $A_1 = A_1^\dagger$, we need

$$2iH(t)dt = 0 + O(dt^2) \implies H(t) = 0. \tag{3.44}$$

From equation 3.35,

$$\begin{aligned} I - P(t)dt + O(dt^2) &= \frac{1}{2}I + \frac{1}{2}(I - \sum_{k\neq 1} A_k^\dagger(t, dt)A_k(t, dt)) + O(dt^2) \\ &= \frac{1}{2}(I + A_1^\dagger(t, dt)A_1(t, dt)) + O(dt^2), \end{aligned} \tag{3.45}$$

which is positive semi-definite (up to order dt^2). So for instantaneous Kraus operators, we have

$$\exp\left(-iH(t)dt - P(t)dt\right) \geq 0 \Leftrightarrow H(t) = 0. \tag{3.46}$$

Thus our notion of a decoherent channel occurs iff $H(t) = 0$ for all t, in which case we would have

$$\frac{d\rho}{dt} = -\frac{1}{2}\left\{\sum_{k\neq 1} L_k^\dagger(t)L_k(t), \rho(t)\right\} + \sum_{k\neq 1} L_k(t)\rho(t)L_k^\dagger(t). \tag{3.47}$$

i.e. just the decoherent part of the Lindblad equation, as interpreted by previous literature. Moreover, if we use the leading Kraus approximaton, the last term drops out and we have

$$\frac{d\rho}{dt} = -\{P(t), \rho(t)\}. \tag{3.48}$$

Such a simple equation can be used to generate decoherent evolution.

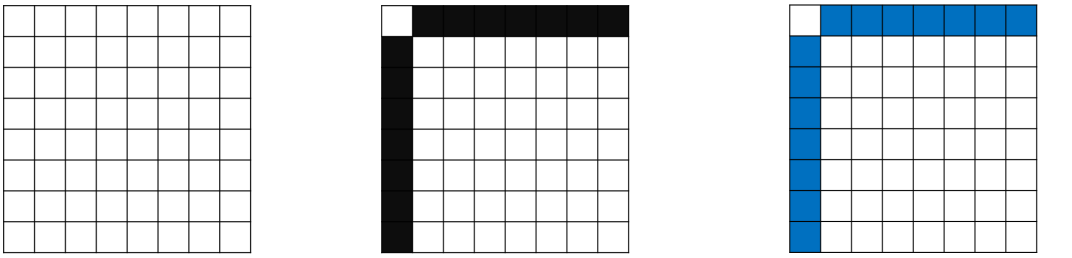
Thus our definition of decoherent channels harmonizes well with notions already established in the literature, at least in the limit of instantaneous operators. As we carry on, we will find that our definition is reasonable even outside of this limit.

3.5 Bitter Channels

We introduced the non-catastrophic condition in section 3.1 in order to remove pathological behaviour from our quantum channels. This condition allowed us to derive several theorems, showcasing how leading Kraus operators capture the fidelity and unitarity of channels and their compositions. Unfortunately, this condition is still loose enough to allow several problematic noise scenarios. We shall explore two of these scenarios here, and implement a method to avoid them.

In realistic noise scenarios, we expect there to be a fair degree of homogeneity in the action of the noise. That is to say, we do not expect a realistic noise process to treat one state of our system extremely poorly, while leaving everything else untouched. N.B. this is a very different statement from requiring that our channels act homogeneously on *qubits*. Adding or removing a single qubit from a system changes the dimension by a factor of 2. A channel focused on one qubit, leaving everything else untouched, could easily influence "half" of the states (in the sense of influencing half of the states in a basis, revealing an influence on half of the dimensions of the space). For instance, for an N qubit system, consider the basis $\{|x_1\rangle \otimes \ldots \otimes |x_N\rangle\}$, where

$$|x_i\rangle = |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ or } |x_i\rangle = |2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \forall i. \tag{3.49}$$



(a) ρ.

(b) $\mathcal{A}(\rho)$.
$A_1 = \sum_{i\neq 1} |i\rangle\langle i|$.

(c) $\mathcal{U}(\rho)$.
$U = \sum_{i\neq 1} |i\rangle\langle i| - |1\rangle\langle 1|$.

Figure 3.1: A visualization of the action of bitter channels on a density matrix. The black and blue squares represent entries of the density matrix, ρ, being sent to 0 and being shifted by a phase of -1, respectively.

We can consider the unitary

$$U = Z \otimes I \otimes \ldots \otimes I \; ; \; Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.50}$$

acting as Z only on the first qubit. If we act the channel $\mathcal{U}$ on our basis, then anything of the form $|1\, x_2 \ldots x_N\rangle$, will be left untouched, while the states $|2\, x_2 \ldots x_N\rangle$ will pick up a minus sign.

In this way, the channel $\mathcal{U}$ is influencing half of the space of pure states, while only affecting one qubit. This is a far cry from the issue that we are confronting: a channel acting poorly on one basis state out of d many. Such a behaviour — precise and antagonistic — would be spiteful in a person; revealing a grudge. We will call such things **bitter channels**.

As an example, consider a leading Kraus operator, $A_1 = \sum_{i\neq 1} |i\rangle\langle i|$. We can fill out the channel with a second Kraus operator, $A_2 = |1\rangle\langle 1|$. Calculating the fidelity, we have $\Phi(\mathcal{A}, \mathcal{I}) = \Upsilon(\mathcal{A}) = (|\mathrm{Tr}(A_1)|^2 + |\mathrm{Tr}(A_2)|^2)/d^2 = ((d-1)^2+1)/d^2 = (1 - 2/d + 2/d^2)$. For a large enough system, such a channel will be non-catastrophic. Yet this is a very unrealistic noise scenario, with $\mathcal{A}(\rho) = \sum_{ij\neq 1} \rho_{ij} |i\rangle\langle j| + \rho_{11} |1\rangle\langle 1|$, completely killing off the coherence between the state $|1\rangle\langle 1|$ and every other (see figure 3.1b). Similarly, consider a unitary channel, $\mathcal{U}$, with Kraus operator $U = \sum_{i\neq 1} |i\rangle\langle i| - |1\rangle\langle 1|$. All unitary channels are non-catastrophic. However based on physical intuition, we might choose to discount this state — it tunes the phase of one canonical basis state, while leaving the others alone (figure 3.1c). This type of manipulation could be achieved intentionally, but it is unlikely to be found occurring in background noise.

The way we eliminate such scenarios is by restricting the variance of the eigenvalues of the operators. As we saw in section 1.3, every normal matrix gives rise to an orthonormal basis of eigenvectors. Let A be normal with eigenvectors $\{|x_i\rangle\}$. We can write the action of A on any state as $A|y\rangle = \sum_i y_i A|x_i\rangle = \sum_i y_i \lambda_i |x_i\rangle$. To achieve relative homogeneity of the action of the operator, we enforce that homogeneity on the eigenvalues. If we enforce too much homogeneity, we will restrict ourselves to only studying the identity channel. Luckily, very little homogeneity will be necessary for our purposes. With that said, let us define our bitter channels.

3.5.1 Bitter Dephasers

Definition 21 (Bitter Dephasers)
Let $\mathcal{D}$ be a non-catastrophic decoherent channel, with canonical Kraus operators $\{M_i\}$. Consider the eigenvalues of the leading Kraus operator of $\mathcal{D}$: $\{\lambda_i(M_1)\}$. $\mathcal{D}$ is called a ***bitter dephaser*** *if one of these eigenvalues deviates from the average much more than the average deviates from one:*

$$1 - \mathbb{E}_i[\lambda_i(M_1)] = 1 - \frac{\mathrm{Tr}(M_1)}{d} \ll |\mathbb{E}_i\left[\lambda_i(M_1)\right] - \lambda_j(M_1)|\,, \tag{3.51}$$

for some λ_j.

We should clarify the sense in which we intend to use "$\ll$" here. Consider the trace preservation condition: $\sum_i A_i^\dagger A_i = I$. Let $|x\rangle$ be a normalized eigenvector of $A_j^\dagger A_j$, with corresponding eigenvalue λ. Then

$$\langle x|\sum_i A_i^\dagger A_i|x\rangle = \lambda + (\text{terms } \geq 0) = 1. \tag{3.52}$$

So every eigenvalue of each $A_j^\dagger A_j$ must be ≤ 1. So the square roots of the these eigenvalues (the singular values of every A_j) are between 0 and 1. It follows that the expectation value of the singular values of any Kraus operator will be between 0 and 1. The quantities appearing in equation 3.51 will then be between 0 and 1 as well (since we are considering the leading Kraus operator of a decoherent channel in definition 21, it will be positive-semidefinite, and thus its singular values and eigenvalues are the same). As such, we say $x \ll y$ in the sense of equation 3.51 iff $y/x \gg 0$ in an absolute sense; e.g. $y/x > 10$.

What is the intuition of definition 21? Using the leading Kraus approximation, we can relate the average of the eigenvalues of the leading Kraus operator to the fidelity of the

channel:

$$\Phi(\mathcal{D},\mathcal{I}) \approx \Phi(\mathcal{D}^\circ,\mathcal{I}) = \frac{|\mathrm{Tr}(M_1)|^2}{d^2} = \left(\frac{\mathrm{Tr}(M_1)}{d}\right)^2 = \mathbb{E}_i[\lambda_i(M_1)]^2. \tag{3.53}$$

So the left hand side of equation 3.51 relates to the infidelity of the channel (though it is not quite the same thing). Since the eigenvalues of matrices tell us about those matrices' actions on vectors, the right hand side of equation 3.51 talks about the inhomogeneity of the action of the channel. Definition 21 thus essentially asks whether the inhomogeneity is much worse than the infidelity, and if so, it labels the channel "bitter".

For those channels with low fidelities, the left hand side of equation 3.51 will be closer to 1. Since the right hand side is bounded above by 1, it becomes less likely for the channel to be a bitter dephaser. For instance, if the fidelity is 0.5, the left hand side of equation 3.51 becomes $1-\sqrt{0.5} \approx 0.29$. Thus for the channel to be bitter, we would need something like

$$\frac{|\mathbb{E}_i\left[\lambda_i(M_1)\right] - \lambda_j(M_1)|}{1-\mathbb{E}_i\left[\lambda_i(M_1)\right]} \approx 3.4\,|\mathbb{E}_i\left[\lambda_i(M_1)\right] - \lambda_j(M_1)| > 10, \tag{3.54}$$

requiring a large discrepancy between λ_j and the average. In fact, since all of the eigenvalues of the leading Kraus operator are between 0 and 1, it's impossible to even satisfy this inequality. One might choose to consider "$\ll$" in equation 3.51 to correspond to, say $y/x > 3$, as opposed to $y/x > 10$, allowing these low-fidelity bitter dephasers, but in any case, we see that lower-fidelity channels will rarely be bitter.

On the other hand, for a fidelity of 0.95, we would need

$$39.5\,|\mathbb{E}_i\left[\lambda_i(M_1)\right] - \lambda_j(M_1)| > 10, \tag{3.55}$$

requiring a much lower inhomogeneity in the action of the channel for the channel to be considered bitter. As a mnemonic, we say we **expect more** from higher fidelity channels, and so we are quicker to call them bitter.

Consider again $A_1 = \sum_{i\neq 1} |i\rangle\langle i|$. It is Hermitian with singular values $0, 1, \ldots, 1$. Examining 3.51, we have

$$\frac{1}{d} = 1 - \frac{d-1}{d} \ll \frac{d-1}{d} - 0 = 1 - \frac{1}{d}, \tag{3.56}$$

since $d-1 > 10$ for $N \geq 4$ ($d = 2^N$). Thus the channel is a bitter dephaser for even small system sizes.

3.5.2 Bitter Unitaries

We now consider the unitary case. We would like an analogous definition to that of definition 21. The entanglement fidelity involves the *absolute value* of the trace of products of Kraus operators. Thus we are free to add phases of our liking onto Kraus operators when we only care about the entanglement fidelity. In particular, we can choose a phase such that the trace of U becomes real and non-negative: if $\mathrm{Tr}(U) = re^{i\phi}$ then consider $e^{-i\phi}U$. Or, in other words,

Definition 22
Let U be a unitary matrix. Define the corresponding ***mean-semi-positive*** *(MSP) unitary, $U^{\curvearrowright}$, to be*

$$U^{\curvearrowright} \equiv \frac{\mathrm{Tr}(U)^*}{|\mathrm{Tr}(U)|}U. \tag{3.57}$$

Note: The notation $\curvearrowright$ is meant to invoke the idea of turning the trace of U (as a complex number) onto the non-negative real axis. A mean-semi-positive unitary should not be confused with a positive semi-definite unitary; since all of the eigenvalues of a unitary matrix have norm 1, a PSD unitary would have all eigenvalues equal to 1, and thus, from the spectral decomposition, $U = VIV^\dagger = I$. So the only PSD unitary is the identity. On the other hand, all of the Pauli matrices are mean-semi-positive.

Since the trace of a matrix is the sum of its eigenvalues, and thus the sum of the real and imaginary parts of each eigenvalue, the imaginary parts of the eigenvalues of a mean-semi-positive matrix must go to zero when averaged out: $\mathrm{Tr}(U)^{\curvearrowright} = \sum_i \mathbb{R}(\lambda_i^{\curvearrowright}) + i\sum_i \mathbb{I}(\lambda_i^{\curvearrowright}) \in \mathbb{R} \implies \sum_i \mathbb{I}(\lambda_i^{\curvearrowright}) = 0$. When we consider the distance between an eigenvalue and the average of $U^{\curvearrowright}$, in the definition of bitter unitaries, we will only be concerned with the real parts of the eigenvalues, because their imaginary parts do not contribute to the average.

Definition 23 (Bitter Unitaries)
Consider a unitary channel, $\mathcal{U}$, with corresponding unitary U, and mean-semi-positive counterpart $U^{\curvearrowright}$. Let $\{\lambda_i^{\curvearrowright}\}$ be the eigenvalues of $U^{\curvearrowright}$. We call $\mathcal{U}$ ***bitter****, if*

$$1 - \mathbb{E}_i[\mathbb{R}(\lambda_i^{\curvearrowright})] = 1 - \frac{\mathrm{Tr}(U^{\curvearrowright})}{d} \ll \left|\mathbb{E}_i[\mathbb{R}(\lambda_i^{\curvearrowright})] - \mathbb{R}(\lambda_j^{\curvearrowright})\right|, \tag{3.58}$$

for some $\lambda_j^{\curvearrowright}$.

We use "$\ll$" in the same sense as for the bitter dephaser case, since the eigenvalues of unitary matrices will sit on the unit circle and thus their real parts will have magnitude ≤ 1.

Consider the polar decomposition, $A_1 = U|A_1|$, of the leading Kraus operator of $\mathcal{A}$. We can apply our pair of bitter conditions to the two components of the decomposition of A_1 (correspondingly, to the coherent and decoherent parts of $\mathcal{A}$), to assure good behaviour.

3.5.3 Equablity

Definition 24 (Equable Quantum Channels)
A non-catastrophic quantum channel, $\mathcal{A} = \mathcal{U}\mathcal{D}$, is called ***equable*** *if both $\mathcal{U}$ and $\mathcal{D}$ are not bitter.*

From the last section we see that, by design, equable channels will be free from certain unrealistic behaviours; we built our definitions of bitter channels specifically to eliminate them. It turns out though, that a weaker restriction upon our channels will be sufficient to derive some useful results. We frame our restriction in terms of two constants.

Definition 25 (Wide Sense Equable (WSE) Channels)
Let $\mathcal{A}$ be a non-catastrophic quantum channel with leading Kraus operator A_1, with polar decomposition $A_1 = U|A_1|$.

We define the ***WSE decoherence constant*** *of $\mathcal{A}$, $\gamma_D(\mathcal{A})$, as:*

$$\gamma_D(\mathcal{A}) \equiv \frac{\mathrm{SDev}[\sigma_i(A_1)]}{\mathbb{E}[1 - \sigma_i(A_1)]}, \tag{3.59}$$

where SDev *denotes the standard deviation.*

Similarly, we define the ***WSE coherence constant*** *$\gamma_U(\mathcal{A})$ to be:*

$$\gamma_U(\mathcal{A}) \equiv \frac{\mathrm{SDev}[\mathbb{R}(\lambda_i(U^{\curvearrowright}))]}{\mathbb{E}[1 - \mathbb{R}(\lambda_i(U^{\curvearrowright}))]}. \tag{3.60}$$

$\mathcal{A}$ is said to be ***wide sense equable*** *if*

$$\gamma_D(\mathcal{A}) \ll 1/\sqrt{\mathbb{E}[1 - \sigma_i(\mathcal{A})]}\ , \tag{3.61}$$

and

$$\gamma_U(\mathcal{A}) \ll 1/\sqrt{\mathbb{E}[1 - \mathbb{R}(\lambda_i(\mathcal{A}))]}\ . \tag{3.62}$$

Wide sense equability loosens up the restraints on our channels quite a bit; as the size of our system increases, wilder behaviour is permitted. Even so, we stand to gain quite a bit by envoking WSE. Consider first the unitarity of a composition of channels:

Theorem 3.5.1
Consider m non-catastrophic channels $\mathcal{A}_i$. Then $\Upsilon(\mathcal{A}_{m:1})$ has the following properties:

(Quasi-monotonicity)

$$\Upsilon(\mathcal{A}_{m:1}) \leq \min_i \Upsilon(\mathcal{A}_i) + (1 - \Upsilon(\mathcal{A}_{m:1}))^2.$$

(Quasi-additivity) (3.63)

$$1 - \Upsilon(\mathcal{A}_{m:1}) \leq \sum_{i=1}^{m} (1 - \Upsilon(\mathcal{A}_i)) + 2\sum_{i=1}^{m} (1 - \Upsilon(\mathcal{A}_i))^2 + \text{ H.O.}$$

These inequalities are almost saturated by bitter channels. If we introduce the WSE decoherence constants (letting γ_D be the largest), we obtain:

$$\begin{aligned}\left| \Upsilon(\mathcal{A}_{m:1}) - \prod_i \Upsilon(\mathcal{A}_i) \right| \leq & \sum_{i=1}^{m} (1 - \Upsilon(\mathcal{A}_i))^2 + (1 - \Upsilon(\mathcal{A}_{m:1}))^2 \\ & + \frac{4\gamma_D^2}{\prod_i \sqrt{\Upsilon(\mathcal{A}_i^\circ)}} \left(1 - \prod_i \sqrt{\Upsilon(\mathcal{A}_i^\circ)} \right)^2 + \text{H.O. ,}\end{aligned} \tag{3.64}$$

*which guarantees **quasi-multiplicativity** of Υ when the errors are WSE.*

Let us take apart each of these three results, one at a time. Quasi-monotonicity tells us that when we compose a series of (non-catastrophic) channels together, the unitarity will be (approximately) bounded by the unitarity of the weakest channel. Considering the unitarity to be a measure of certainty in our operations, we say that we can only be as confident as we are in our worst channel.

In viewing quasi-subadditivity, we note that if $\Upsilon(\mathcal{A})$ describes the coherent behaviour of a channel, then $1 - \Upsilon(\mathcal{A})$ is a measure of the incoherent. Quasi-additivity tells us that the incoherent behaviour of a composition of non-catastrophic channels will not expand much faster than linearly. Thus a reasonable amount of coherent behaviour will be preserved through each composition.

In the case of WSE channels, quasi-multiplicity points to a level of independence between the channels. The unitarity of a composition is (roughly) independent of the order of the channels composed, and there is no significant "cross term" denoting channel interactions. Since the entanglement fidelity is always between 0 and 1, we have that $\prod_i \Upsilon(\mathcal{A}_i) \leq \min(\Upsilon(\mathcal{A}_i))$. Thus quasi-multiplicativity is a tighter bound than quasi-monotonicity.

Recalling the relationship between $\Upsilon(\mathcal{A})$ and the unitarity in equation 2.34, it follows that

$$u(\mathcal{A}_{m:1}) \approx \frac{d^2 \prod_{i=1}^{m} \Upsilon(\mathcal{A}_i) - 1}{d^2 - 1} \tag{3.65}$$

for WSE channels.

Theorem 3.5.2
Consider a non-catastrophic channel, $\mathcal{A}$, with polar decomposition $\mathcal{A} = \mathcal{V}\mathcal{D}$ and unitary target $\mathcal{U}$. Then, the maximal unitary correction of $\mathcal{A}$ (in terms of Φ) is approximately bounded by the interval $\left[\Upsilon(\mathcal{A}), \sqrt{\Upsilon(\mathcal{A})}\right]$:

$$\begin{aligned} \max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}, \mathcal{U}) &\leq \sqrt{\Upsilon(\mathcal{A})} + (1 - \Upsilon(\mathcal{A}))^2 + \text{H.O.} \ , \\ \max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}, \mathcal{U}) &\geq \Upsilon(\mathcal{A}) - (1 - \Upsilon(\mathcal{A}))^2 \ . \end{aligned} \tag{3.66a}$$

Moreover, if we introduce the WSE decoherence constant, γ_D, we obtain:

$$\max_{W \in SU(d)} \Phi(\mathcal{W} \circ \mathcal{A}, \mathcal{U}) \geq \sqrt{\Upsilon(\mathcal{A})} - (1 - \Upsilon(\mathcal{A}))^2 - \gamma_D^2 \left(1 - \sqrt{\Upsilon(\mathcal{A})}\right)^2 \ . \tag{3.67}$$

If the error attached to $\mathcal{A}$ is WSE, then the interval virtually collapses and $\max_{W \in SU(d)} \Phi(\mathcal{W}\mathcal{A}, \mathcal{U}) \approx \sqrt{\Upsilon(\mathcal{A})}$.

A quasi-maximal choice of unitary correction is $\mathcal{W} = \mathcal{U}\mathcal{V}^\dagger$.

This theorem is our first taste of the idea that all that can be corrected in a channel is the unitary component: $\Phi(\mathcal{W}\mathcal{A}, \mathcal{U}) = \Phi(\mathcal{U}\mathcal{V}^\dagger\mathcal{V}\mathcal{D}, \mathcal{U}) = \Phi(\mathcal{D}, \mathcal{I})$. We will see this theme recapitulated several more times in this work. Moreover, this theorem also tells us that $\Phi(\mathcal{D}, \mathcal{I}) \approx \sqrt{\Upsilon(\mathcal{D})}$ (for WSE channels), relating back to the behaviour of depolarizing channels, just as we would like.

Translating the above result into terms of the average fidelity and unitary, we have

$$\max_{V \in SU(d)} F(\mathcal{V}\mathcal{A}, \mathcal{U}) \approx \frac{\sqrt{(d^2 - 1)u(\mathcal{A}) + 1} + 1}{d + 1} \ . \tag{3.68}$$

We next apply WSE to investigate decoherent channels.

3.6 WSE and Decoherent Channels

Theorem 3.6.1
Consider m non-catastrophic decoherent channels $\mathcal{D}_i$ and any non-catastrophic unitary channel $\mathcal{U}$. Then,

(Quasi-monotonicity)

$$\begin{aligned}\Phi(\mathcal{U}\mathcal{D}_{m:1},\mathcal{I}) \leq \min_i \Phi(\mathcal{D}_i,\mathcal{I}) + \frac{1}{2}\left(\sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i^\circ))\right)^2 \\ + (1-\Phi(\mathcal{U}\circ\mathcal{D}_{m:1}^\circ,\mathcal{I}))\sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i^\circ))\,.\end{aligned} \tag{3.69a}$$

(Quasi-subadditivity)

$$\begin{aligned}1-\Phi(\mathcal{U}\mathcal{D}_{m:1},\mathcal{I}) \leq (1-\Phi(\mathcal{U},\mathcal{I})) + \sum_{i=1}^{m}(1-\Phi(\mathcal{D}_i,\mathcal{I})) \\ + (1-\Phi(\mathcal{U},\mathcal{I}))^2 + \sum_{i=1}^{m}(1-\Phi(\mathcal{D}_i^\circ,\mathcal{I}))^2 \\ + \sum_{i=1}^{m}(1-\Phi(\mathcal{D}_i,\mathcal{I}))(1-\Upsilon(\mathcal{D}_i)).\end{aligned} \tag{3.69b}$$

If we enforce the WSE condition we arrive at:

Theorem 3.6.2
Consider m non-catastrophic, decoherent channels $\mathcal{D}_i$ (with target $\mathcal{I}$) with largest WSE decoherence constant, γ_D. Then, $\Phi(\mathcal{D}_{m:1}, \mathcal{I})$ is bounded as follows:

***(Quasi-multiplicativity)**:*

$$\begin{aligned}\left|\Phi(\mathcal{D}_{m:1},\mathcal{I}) - \prod_{i=1}^{m}\Phi(\mathcal{D}_i,\mathcal{I})\right| \leq \Bigg[\frac{1}{2}\left(\sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i^\circ))\right)^2 \\ + (1-\Phi(\mathcal{D}_{m:1}^\circ,\mathcal{I}))\sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i^\circ)) \quad + \sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i))\,(1-\Phi(\mathcal{D}_i,\mathcal{I})) \\ + \frac{\gamma_D^2}{\prod_{i=1}^{m}\sqrt{\Phi(\mathcal{D}_i^\circ,\mathcal{I})}}\left(1-\prod_{i=1}^{m}\sqrt{\Phi(\mathcal{D}_i^\circ,\mathcal{I})}\right)^2\Bigg] + \text{H.O.}\end{aligned} \tag{3.70}$$

If the channels are WSE, $\Phi(\mathcal{D}_{m:1}, \mathcal{I})$ is essentially multiplicative.

We see we have arrived at a set of results analogous to those of theorem 3.5.1, relating to the entanglement fidelity rather than to the unitarity. Moreover, the quasi-multiplicity relationship, $\Phi(\mathcal{D}_{m:1}, I) \approx \prod_i \Phi(\mathcal{D}_i, I)$ harkens back to equation 2.40.

The quasi-monotonicity of our channels lets us know that a composition of (non-catastrophic) decoherent channels cannot be substantially corrected by any unitary channel. Taking this result with theorem 3.5.2, we suspect that there is little coherent buildup in a composition of decoherent channels (analogous to the fact that compositions of depolarizing channels remain depolarizing). The quasi-multiplicativity expresses a level of independence in WSE decoherent channels. From theorem 3.5.1 we were already afforded a level of independence in *any* WSE channel composition. However, that independence was from the unitarity, whereas this second independence is from the fidelity. This implies that decoherent WSE channels may be "even more" independent than others. Again, this makes sense as a generalization of depolarizing channels, since from equation 2.39, we see that depolarizing channels commute, and their unitarities and fidelities are essentially multiplicative.

Using the correspondence between the entanglement fidelity and average process fidelity, we arrive at:

$$F(\mathcal{D}_{m:1},\mathcal{I}) \approx \frac{d\prod_{i=1}^{m}\Phi(\mathcal{D}_i)+1}{d+1}\,. \tag{3.71}$$

3.7 The Coherence Level

Consider, further, the consequences of appending a coherent operation to end of our decoherent composition:

Theorem 3.7.1
Consider m non-catastrophic, decoherent error channels $\mathcal{D}_i$ (with target $\mathcal{I}$). Let γ_D be the largest of their WSE decoherence constants. Moreover, consider a non-catastrophic unitary error channel, $\mathcal{U}$, with WSE coherence constant, γ_U. Then, $\Phi(\mathcal{U}\mathcal{D}_{m:1},\mathcal{I})$ is bounded as follows:

$$\begin{aligned}\left|\Phi(\mathcal{U}\mathcal{D}_{m:1},\mathcal{I}) - \Phi(\mathcal{U},\mathcal{I})\prod_{i=1}^{m}\Phi(\mathcal{D}_i,\mathcal{I})\right| \leq \Bigg[&\frac{1}{2}\left(\sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i^\circ))\right)^2 \\ &+ (1-\Phi(\mathcal{U}\mathcal{D}_{m:1}^\circ,\mathcal{I}))\sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i^\circ)) + \sum_{i=1}^{m}(1-\Upsilon(\mathcal{D}_i))\,(1-\Phi(\mathcal{D}_i,\mathcal{I})) \\ &+ \frac{\gamma_D^2}{\prod_{i=1}^{m}\Phi(\mathcal{D}_i^\circ,\mathcal{I})}\left(1-\prod_{i=1}^{m}\sqrt{\Phi(\mathcal{D}_i^\circ,\mathcal{I})}\right)^2 \\ &+ 2\frac{\gamma_D\gamma_U}{\prod_{i=1}^{m}\sqrt{\Phi(D_i^\circ,I)}}\left(1-\prod_{i=1}^{m}\sqrt{\Phi(\mathcal{D}_i^\circ,\mathcal{I})}\right)\left(1-\sqrt{\Phi(\mathcal{U},\mathcal{I})}\right)\Bigg] + \text{H.O.}\end{aligned} \tag{3.72}$$

If the errors are WSE, then $\Phi(\mathcal{U}\mathcal{D}_{m:1},\mathcal{I})$ is essentially multiplicative:
$\Phi(\mathcal{U}\mathcal{D}_{m:1},\mathcal{I}) \approx \Phi(\mathcal{U},\mathcal{I})\prod_i \Phi(\mathcal{D}_i,\mathcal{I})$.

Consider the case $m = 1$. Let $\mathcal{A}$ be a channel with target $\mathcal{U}$ and polar decomposition $\mathcal{V}\mathcal{D}$. $\mathcal{W} := \mathcal{U}^\dagger\mathcal{V}$ is a unitary channel. Thus by theorem 3.7.1 we have

$$\Phi(\mathcal{A},\mathcal{U}) = \Phi(\mathcal{W}\mathcal{D},\mathcal{I}) \approx \Phi(\mathcal{W},\mathcal{I})\Phi(\mathcal{D},\mathcal{I}) = \Phi(\mathcal{V},\mathcal{U})\Phi(\mathcal{D},\mathcal{I}) \approx \Phi(\mathcal{V},\mathcal{U})\sqrt{\Upsilon(\mathcal{A})}\ . \tag{3.73}$$

Breaking this down, we can approximate the entanglement fidelity as being composed of two parts: $\Phi(\mathcal{V},\mathcal{U})$, a coherent contribution, from the unitary part of the channel, $\mathcal{A}$, and $\Phi(\mathcal{D},\mathcal{I}) \approx \sqrt{\Upsilon(\mathcal{A})}$, a decoherent one. Again, we see a level of independence appearing between channels; this time between the coherent and decoherent parts of $\mathcal{A}$. Keeping in mind the independence between (WSE) decoherent channels, the full result of 3.7.1 (for a general composition length) is almost expected.

At this point, we shall reformulate some of our results in terms of the **infidelity** ($r(\mathcal{A},\mathcal{U}) \equiv 1 - F(\mathcal{A},\mathcal{U})$), as such language is often used in randomized benchmarking literature. Translating into terms of the infidelity gives us (up to $O(r^2)$):

$$r(\mathcal{A},\mathcal{U}) \approx \underbrace{r(\mathcal{V},\mathcal{U})}_{\text{Coherent infidelity}} + \underbrace{r(\mathcal{D},\mathcal{I})}_{\text{Decoherent infidelity}} \equiv r_{\text{coh}} + r_{\text{decoh}}, \tag{3.74}$$

where, by theorem 3.6.1, r_{decoh} is not substantially correctable by any composition, and can be computed through the square root of the unitarity (theorem 3.5.2). We have

$$r_{\text{decoh}} = \frac{d - \sqrt{(d^2-1)u(\mathcal{A}) + 1}}{d+1} + O(r^2). \tag{3.75}$$

On the other hand, r_{coh} can be corrected through the proper unitary inversion.

Using the above decomposition, we can easily define the level of coherent behaviour in a (WSE) channel:

Definition 26 (Coherence Level)
*Let $\mathcal{A}$ be a non-catastrophic WSE quantum channel. The **coherence level** of $\mathcal{A}$ is*

$$\frac{r_{\text{coh}}}{r} \equiv 1 - \frac{d - \sqrt{(d^2-1)u(\mathcal{A}) + 1}}{(d+1)r(\mathcal{A},\mathcal{U})} + O(r) \ . \tag{3.76}$$

*The **decoherence level** of $\mathcal{A}$ is*

$$\frac{r_{\text{decoh}}}{r} \equiv 1 - \frac{r_{\text{coh}}}{r} \ . \tag{3.77}$$

A similar notion of decoherence level was presented in [13, 32] under the names "incoherent error" and "incoherence", respectively. In those works, r_{decoh} is shown to be the smallest infidelity achievable through unitary correction. In [13], equation 3.74 was derived, but only for the case of one qubit. Our results extend those previously found and place them within a new context of WSE channels.

Next, consider theorem 3.7.1 for a general circuit depth. We shall find that we can provide bounds on the best and worst case fidelity of a correction.

3.8 Worst and Best Case Fidelity Correction

Consider m channels, $\mathcal{A}_i$, with targets, $\mathcal{U}_i$, and polar decompositions, $\mathcal{D}_i\mathcal{V}_i$. By using these decompositions and inserting identity channels appropriately into the composition, we can

express $\mathcal{A}_{m:1}$ as:

$$\mathcal{A}_{m:1} = \mathcal{V}_{m:1}(\mathcal{V}_{m:1})^\dagger \mathcal{D}_m \mathcal{V}_{m:1} \ldots (\mathcal{V}_{2:1})\mathcal{V}_1^\dagger \mathcal{D}_1 \mathcal{V}_1 = \mathcal{V}_{m:1}\mathcal{D}'_{m:1} \ , \tag{3.78}$$

with $\mathcal{D}'_k := (\mathcal{V}_{k:1})^\dagger \mathcal{D}_k \mathcal{V}_{k:1}$. Each $\mathcal{D}'_k$ has leading Kraus operator, $V_{k:1}A_{k,1}V_{k:1}^\dagger$, where $V_{k:1} = V_k \ldots V_1$ and $A_{k,1}$ is the leading Kraus operator of $\mathcal{D}_k$. The leading Kraus operator of each $\mathcal{D}'_k$ is then PSD, and so each $\mathcal{D}'_k$ is decoherent. Furthermore, $\Phi(\mathcal{D}'_k, \mathcal{I}) = \Phi(\mathcal{D}_k, \mathcal{I})$ for all k, by the cylic property of the trace.

This means that:

$$\begin{aligned}
\Phi(\mathcal{A}_{m:1}, \mathcal{U}_{m:1}) &= \Phi(\mathcal{V}_{m:1}\mathcal{D}'_{m:1}, \mathcal{U}_{m:1}) \\
&\approx \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_{i=1}^{m} \Phi(\mathcal{D}'_i, \mathcal{I}) \\
&= \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_{i=1}^{m} \Phi(\mathcal{D}_i, \mathcal{I}) \\
&\approx \Phi(\mathcal{V}_{m:1}, \mathcal{U}_{m:1}) \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)} \ .
\end{aligned} \tag{3.79}$$

And so we see that the fidelity of *any* composition of realistic channels can be factored into coherent and decoherent contributions. Knowing that we can only correct the coherent part of the composition, we arrive at:

Theorem 3.8.1
Consider m non-catastrophic channels $\mathcal{A}_i$ with respective unitary targets $\mathcal{U}_i$ and polar decompositions $\mathcal{A}_i = \mathcal{V}_i\mathcal{D}_i$. Let the largest WSE decoherence constant be γ_D. Then the maximal unitary correction of the composition, $\mathcal{A}_{m:1}$, is bounded as follows:

$$\max_{W\in SU(d)} \Phi(\mathcal{W}\mathcal{A}_{m:1},\mathcal{U}_{m:1}) - \prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)} \leq \left[\frac{5}{4}\left(\sum_{i=1}^{m}(1-\Upsilon(\mathcal{A}_i))\right)^2 + \sum_{i=1}^{m}(1-\Upsilon(\mathcal{A}_i))^2 + \left(\sum_{i=1}^{m}(1-\Upsilon(\mathcal{A}_i))\right)\left(1-\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)}\right) + \frac{2\gamma_D^2}{\prod_{i=1}^{m}\Upsilon(A_i^\circ)}\left(1-\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i^\circ)}\right)^2\right] + \text{H.O.} \tag{3.80a}$$

$$\max_{W\in SU(d)} \Phi(\mathcal{W}\mathcal{A}_{m:1},\mathcal{U}_{m:1}) - \prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)} \geq \left[-\gamma_D^2\sum_{i=1}^{m}\left(1-\sqrt{\Upsilon(A_i^\circ)}\right)^2 - \sum_{i=1}^{m}(1-\Upsilon(A_i))^2 - \frac{\gamma_D^2}{\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i^\circ)}}\left(1-\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i^\circ)}\right)^2\right] + \text{H.O.} \tag{3.80b}$$

For WSE errors, the maximal unitary correction of the composition, $\mathcal{A}_{m:1}$, is essentially $\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)}$. A quasi-optimal choice of unitary correction is $\mathcal{W} = \mathcal{U}_{m:1}(\mathcal{V}_{m:1})^\dagger$.

Or, in terms of the average gate fidelity,

$$\max_{W\in SU(d)} F(\mathcal{W}\circ\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \approx \frac{d\prod_{i=1}^{m}\Phi(\mathcal{D}_i,\mathcal{I})+1}{d+1} \approx \frac{d\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)}+1}{d+1}\,. \tag{3.81}$$

We are also in a position to determine the worst possible corrected fidelity. In [4] the following inequality was found:

$$\Phi(\mathcal{V}_{m:1},\mathcal{U}_{m:1}) \geq \cos^2\left(\sum_{i=1}^{m}\arccos\left(\sqrt{\Phi(\mathcal{V}_i,\mathcal{U}_i)}\right)\right). \tag{3.82}$$

This inequality is saturated in even dimensions, by commuting unitary errors of the form

$$\begin{pmatrix}\cos(\theta_i) & -\sin(\theta_i)\\ \sin(\theta_i) & \cos(\theta_i)\end{pmatrix}\otimes I_{d/2}. \tag{3.83}$$

For the odd dimensional case, we have the saturated bound

$$\Phi(\mathcal{V}_{m:1},\mathcal{U}_{m:1}) \geq \left(\frac{(d-1)\cos\left(\sum_{i=1}^{m}\arccos\left(\frac{d\sqrt{\Phi(\mathcal{V}_i,\mathcal{U}_i)}-1}{d-1}\right)\right)+1}{d} \right)^2 , \tag{3.84}$$

computed using the result for the even dimensional case and applying the phase freedom in $\Phi(\mathcal{A},\mathcal{B})$ to fix an eigenvalue and reduce the dimension.

By using $\Phi(\mathcal{V}_i,\mathcal{U}_i) \approx \Phi(\mathcal{A}_i,\mathcal{U}_i)/\sqrt{\Upsilon(\mathcal{A}_i)}$ with the above result, we can arrive at quasi-bounds for the fidelity (in the WSE regime):

For even dimensions:

$$\cos^2\left(\sum_{i=1}^{m}\arccos\left(\sqrt{\frac{\Phi(\mathcal{A}_i,\mathcal{U}_i)}{\sqrt{\Upsilon(\mathcal{A}_i)}}}\right)\right)\prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)} \lessapprox \Phi(\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \lessapprox \prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)} \; ; \tag{3.85a}$$

For odd dimensions:

$$\left(\frac{(d-1)\cos\left(\sum_{i=1}^{m}\arccos\left(\frac{d\sqrt{\frac{\Phi(\mathcal{A}_i,\mathcal{U}_i)}{\sqrt{\Upsilon(\mathcal{A}_i)}}}-1}{d-1}\right)\right)+1}{d} \right)^2 \prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)} \lessapprox \Phi(\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \lessapprox \prod_{i=1}^{m}\sqrt{\Upsilon(\mathcal{A}_i)} \; . \tag{3.85b}$$

The sum of arccos terms is similar to the so-called **coherence angles** investigated in [4]. Their sum can be interpreted as a coherent buildup. In the case of a decoherent composition, we expect no coherent buildup, and indeed from theorem 3.8.1 we see that the sum of arccos terms go to 0 (mod 2π).

3.9 Decoherence-limited Channels

We have already explored the behaviour of compositions of purely decoherent channels, culminating in theorem 3.8.1. The behaviour that we found should be preserved if we add in a sufficiently small amount of coherence to our channels.

Definition 27 (Decoherence limited channels)
Let $\mathcal{A}$ be a non-catastrophic WSE channel, such that

$$\Phi(\mathcal{A},\mathcal{U}) = \sqrt{\Upsilon(\mathcal{A})} + O(r(\mathcal{A},\mathcal{U})^2), \tag{3.86}$$

for some target, $\mathcal{U}$. Then $\mathcal{A}$ is called a ***decoherence-limited*** *channel.*

In particular, any channel with sufficiently small coherence level — $r_{\text{coh}} = O(r(\mathcal{A},\mathcal{U})^2)$ — will be decoherence-limited, since equation 3.74 holds up to order $r(\mathcal{A},\mathcal{U})^2$, and so

$$r(\mathcal{A},\mathcal{U}) = r_{\text{decoh}} + r_{\text{coh}} + O(r(\mathcal{A},\mathcal{U})^2) = r_{\text{decoh}} + O(r(\mathcal{A},\mathcal{U})^2). \tag{3.87}$$

Translating back into the fidelity gives equation 3.86.

Consider a sequence of decoherence-limited channels, $\{\mathcal{A}_i\}_{i=1}^m$. Since we have only a small amount of coherence, the arguments in the cosine terms in equation 3.85 must be small. We can thus Taylor expand the cosine:

$$\cos^2(x) = (1 - \frac{x^2}{2} + \ldots)^2 \approx 1 - x^2. \tag{3.88}$$

So equation 3.85 reduces to

$$(1-x^2)\prod_{i=1}^m \sqrt{\Upsilon(\mathcal{A}_i)} \lessapprox \Phi(\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \lessapprox \prod_{i=1}^m \sqrt{\Upsilon(\mathcal{A}_i)}, \tag{3.89}$$

where x is the appropriate argument in the even and odd dimensional cases. Since each channel is decoherence-limited, we have

$$\frac{\Phi(\mathcal{A}_i,\mathcal{U}_i)}{\sqrt{\Upsilon(\mathcal{A}_i)}} = 1 + \frac{O(r(\mathcal{A}_i,\mathcal{U}_i)^2)}{\sqrt{\Upsilon(\mathcal{A}_i)}}. \tag{3.90}$$

However, we are working with non-catastrophic channels, giving us $\Upsilon(\mathcal{A}_i) > 1/2$. Thus the $O(r^2)/\sqrt{\Upsilon(\mathcal{A}_i)}$ above will never be larger than $\sqrt{2}O(r^2)$, and so we consider this to be $O(r^2)$.

Next we note that

$$\sqrt{1+x^2} = 1 + \frac{x^2}{2} + \ldots \tag{3.91}$$

And so

$$\sqrt{\frac{\Phi(\mathcal{A}_i,\mathcal{U}_i)}{\sqrt{\Upsilon(\mathcal{A}_i)}}} = 1 + \frac{1}{2}O(r(\mathcal{A}_i,\mathcal{U}_i)^2) = 1 + O(r(\mathcal{A}_i,\mathcal{U}_i)^2). \tag{3.92}$$

Finally, expanding arccosine, we have

$$\arccos(1-x) = \sqrt{2x} + \frac{(2x)^{3/2}}{24} + \ldots \; ; \; x \geq 0. \tag{3.93}$$

And so the sum over arccosine terms reduces to

$$\sum_{i=1}^{m} \arccos\left(\sqrt{\frac{\Phi(\mathcal{A}_i,\mathcal{U}_i)}{\sqrt{\Upsilon(\mathcal{A}_i)}}}\right) = \sum_{i=1}^{m} O(r(\mathcal{A}_i,\mathcal{U}_i)). \tag{3.94}$$

A similar analysis holds for the argument of the arccosine term in the odd dimensional case. Thus equation 3.89 is transformed into

$$(1 - \left(\sum_i O(r(\mathcal{A}_i,\mathcal{U}_i))\right)^2) \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)} \lessapprox \Phi(\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \lessapprox \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)}. \tag{3.95}$$

We can express the infidelities in terms of their coherent and decoherent components. Since each of our channels is decoherence-limited, the coherent contribution is order r^2, so we can see the above sum, $\sum_i O(r_i)$, as a sum of decoherent terms, $\sum_i O(r_{\text{decoh},i})$. Then using the fact that decoherent fidelities factor (theorem 3.6.2) and $1-\sum_i r_i \approx (1-r_1)\ldots(1-r_m)$ for small r, we have that

$$\sum_i O(r_{\text{decoh},i}) = O(1 - \Phi_{\text{decoh},1}\ldots\Phi_{\text{decoh},m}) = O(1 - \Phi_{\text{decoh},m:1}) = O(r_{\text{decoh},m:1}). \tag{3.96}$$

Thus for WSE channels, $\sum_i O(r(\mathcal{A}_i,\mathcal{U}_i)) = O(r_{\text{decoh}}(\mathcal{A}_{m:1},\mathcal{U}_{m:1}))$. So we have

$$(1 - O(r_{\text{decoh}}(\mathcal{A}_{m:1},\mathcal{U}_{m:1})^2) \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)} \lessapprox \Phi(\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \lessapprox \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)}. \tag{3.97}$$

$0 \leq \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)} \leq 1, \; \forall m$, so it can be absorbed into the order. Also since $r \approx r_{\text{decoh}} + r_{\text{coh}}$, $O(r_{\text{decoh}}) = O(r)$. Since the inequalities hold up to order r^2, we must have

$$\prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)} + O(r(\mathcal{A}_{m:1},\mathcal{U}_{m:1})^2) \leq \Phi(\mathcal{A}_{m:1},\mathcal{U}_{m:1}) \leq \prod_{i=1}^{m} \sqrt{\Upsilon(\mathcal{A}_i)} + O(r(\mathcal{A}_{m:1},\mathcal{U}_{m:1})^2) \tag{3.98}$$

And so the coherent behaviour of the composition is bounded to be order r^2 — the composition is decoherence limited.

This result allows us passage into studying the structure of a subset of decoherent channels. In general, decoherent channels are not preserved under composition: there can be some level of coherent behaviour which develops as they are composed. However, restricting our attention to WSE decoherence-limited channels, we are afforded closure. This is reminiscent of the closure of the set of depolarizing channels.

Chapter 4

Conclusions and Future Directions

In this work we have developed the leading Kraus approximation for quantum channels. Though this approximation generally produces non-physical channels, we have shown that for a particular class of channels (non-catastrophic), the approximate channel will provide accurate estimations for the fidelity and unitarity of the full channel. We have also introduced the notion of wide sense equability, and have argued that realistic noise scenarios will satisfy both the non-catastrophic and wide sense equable conditions.

The leading Kraus approximation led us to a simple decomposition of quantum channels into coherent and decoherent components. We proceeded to investigate decoherent and coherent channels under non-catastrophic and wide sense equable conditions. In particular, it was found that decoherent channels, as we have defined them, fit in well with ideas well-established in the literature, and are shown to generalize results for depolarizing channels: the quasi-multiplicativity of their fidelities, their uncorrectability, and the quadratic relationship between their fidelities and unitarities. Defining "decoherence-limited" channels, we were able to generalize the closure under composition of depolarizing channels to this subset of decoherent channels.

The leading Kraus approximation declutters the Kraus representations of quantum channels, and makes investigations into them more intuitive; removing superfluous information, to uncover approaches which may have been obscured when viewing the channels in their full scope. This technique can be a springboard for developing new results. It can also be useful in revisiting old work. We can apply the LKA in combination with well-established protocols and see if the simplifications it allows (for realistic channels) can provide insight into building extensions of those protocols. For instance, because equation 3.79 partitions the average fidelity of a composition of (WSE) channels into coherent and

decoherent contributions, we should be able to use estimates of the decoherence level (obtained from the unitarity) and the full fidelity of a channel, in order to pinpoint its level of coherence. This type of coherence level estimation could be implemented as an extension of a randomized benchmarking protocol.

From more of a structural standpoint, now that we have put in place additional classes of channels — viz. WSE, extremal, decoherent, and decoherence-limited channels — it would worthwhile to explore these forms through topological or algebraic means.

References

[1] M Benito, JR Petta, and Guido Burkard. Optimized cavity-mediated dispersive two-qubit gates between spin qubits. *arXiv preprint arXiv:1902.07649*, 2019.

[2] Sergey Bravyi, Matthias Englbrecht, Robert König, and Nolan Peard. Correcting coherent errors with surface codes. *npj Quantum Information*, 4(1):55, 2018.

[3] Arnaud Carignan-Dugas, Matthew Alexander, and Joseph Emerson. A polar decomposition for quantum channels (with applications to bounding error propagation in quantum circuits). *arXiv e-prints*, page arXiv:1904.08897, Apr 2019.

[4] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. Bounding the average gate fidelity of composite channels using the unitarity. *arXiv e-prints*, page arXiv:1610.05296, Oct 2016.

[5] Hilary A Carteret, Daniel R Terno, and Karol Życzkowski. Dynamics beyond completely positive maps: Some properties and applications. *Physical Review A*, 77(4):042113, 2008.

[6] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285 – 290, 1975.

[7] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin. Practical characterization of quantum devices without tomography. *Phys. Rev. Lett.*, 107:210404, Nov 2011.

[8] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, Jul 2009.

[9] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009.

[10] Simon J Devitt, William J Munro, and Kae Nemoto. Quantum error correction for beginners. *Reports on Progress in Physics*, 76(7):076001, 2013.

[11] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347–S352, Sep 2005.

[12] Richard R Ernst, Geoffrey Bodenhausen, Alexander Wokaun, et al. Principles of nuclear magnetic resonance in one and two dimensions. 1987.

[13] Guanru Feng, Joel J. Wallman, Brandon Buonacorsi, Franklin H. Cho, Daniel K. Park, Tao Xin, Dawei Lu, Jonathan Baugh, and Raymond Laflamme. Estimating the coherence of noise in quantum control of a solid-state qubit. *Phys. Rev. Lett.*, 117:260501, Dec 2016.

[14] Caroline Figgatt, Aaron Ostrander, Norbert M Linke, Kevin A Landsman, Daiwei Zhu, Dmitri Maslov, and Christopher Monroe. Parallel entangling operations on a universal ion-trap quantum computer. *Nature*, page 1, 2019.

[15] Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Phys. Rev. Lett.*, 106:230501, Jun 2011.

[16] Vittorio Gorini, Andrzej Kossakowski, and Ennackal Chandy George Sudarshan. Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17(5):821–825, 1976.

[17] Brian Hall. *Lie groups, Lie algebras, and representations: an elementary introduction*, volume 222. Springer, 2015.

[18] Timothy F Havel. Robust procedures for converting among lindblad, kraus and matrix representations of quantum dynamical semigroups. *Journal of Mathematical Physics*, 44(2):534–557, 2003.

[19] 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, Jan 2008.

[20] Connor Kupchak, Philip J Bustard, Khabat Heshami, Jennifer Erskine, Michael Spanner, Duncan G England, and Benjamin J Sussman. Time-bin-to-polarization conversion of ultrafast photonic qubits. *Physical Review A*, 96(5):053812, 2017.

[21] Goran Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, 1976.

[22] Cecilia C. López, Ariel Bendersky, Juan Pablo Paz, and David G. Cory. Progress toward scalable tomography of quantum maps using twirling-based methods and information hierarchies. *Phys. Rev. A*, 81:062113, Jun 2010.

[23] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:180504, May 2011.

[24] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, Apr 2012.

[25] Ashley Montanaro. Quantum algorithms: an overview. *Npj Quantum Information*, 2:15023, Jan 2016.

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

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

[28] Mary Beth Ruskai, Stanislaw Szarek, and Elisabeth Werner. An analysis of completely-positive trace-preserving maps on m2. *Linear Algebra and its Applications*, 347(1):159 – 187, 2002.

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

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

[31] Joel Wallman, Chris Granade, Robin Harper, and Steven T Flammia. Estimating the coherence of noise. *New Journal of Physics*, 17(11):113020, nov 2015.

[32] CH Yang, KW Chan, R Harper, W Huang, T Evans, JCC Hwang, B Hensen, A Laucht, T Tanttu, FE Hudson, et al. Silicon qubit fidelities approaching incoherent noise limits via pulse engineering. *Nature Electronics*, 2(4):151, 2019.