

Imperial College London

MSci INDIVIDUAL PROJECT

IMPERIAL COLLEGE LONDON

DEPARTMENT OF PHYSICS

Noisy Quantum Process Tomography under varying preparation designs

Supervisor:

Dr. Hyukjoon Kwon

Assessor:

Prof Myungshik Kim

Author:

Irtaza Khalid

Project Code:

QOLS-Kwon-1

CID:

01259694

November 7, 2025

Word Count: 9,997

Abstract

Using a parametrizable preparation design or initial state set, the effects of preparation on general quantum process tomography are quantitatively explored while keeping the measurement design to be the fixed finite Pauli basis ensemble. A regularized computational model with varying preparation designs is constructed to simulate standard process tomography on n -qubits under decohering noise, which is tweaked to converge. The estimated channels are compared using the completely bounded trace norm or diamond norm and a power-law performance model. Results are obtained for the 1-qubit which has 4 preparation design degrees of freedom. Variation of 2 degrees of freedom explored under the noise model yields no improvement over the standard scheme parametrization. Selective exploration of all 4, however, produces computational and symmetry-based necessary conditions for optimality: preparation designs should be minimal and symmetric. Under these conditions, the tetrahedral design, which contains 4 states, is optimal and outperforms all other designs in all the considered channel tomographies.

Declarations

I declare that this MSci report has been composed by me alone. The work submitted is my own, except where, to the best of my ability, the contributions of others have been included, referenced explicitly and acknowledged. Everything in [Chapter 3](#), apart from some parts which are explicitly highlighted, and [Chapter 4](#) is original and a product of my research.

Acknowledgements

First and foremost, I would like to thank Dr. Hyukjoon Kwon for supervising my project and bearing considerately with my many faults and requests, answering countless questions and suggesting many interesting topics during the course of this project. Our occasional long discussions were valuable and highly enjoyable. Secondly, I would like to thank my parents for letting me take this opportunity to be able to break away from the mould and try something new with my life and being supportive every step of the way.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Thesis overview	3
2	Background	5
2.1	Quantum state representation, postulates and ensembles	5
2.2	Channel representations	8
2.3	Major quantum operations in the geometric picture	11
2.4	Quantum state and process/channel tomography	13
2.5	Channel error and distance metrics	15
2.6	Semi-definite program to compute the diamond norm	16
2.7	Scaling laws for minimax-estimator risk with ensemble size	17
3	Computational standard QPT model	18
3.1	Maximum likelihood method: initial approach and pitfalls	18
3.2	Naturally regularized SQPT model with fixed measurement design .	20
3.3	Noise model	21
3.4	Tests	22
3.5	Power law fit for performance evaluation	23
4	Results: varying preparation designs	25
4.1	Noisy channels	26
4.2	Single axis rotations	27
4.3	Dual axis rotations	28
4.4	Issues with the rotations' approach	31
4.5	Change of basis: selective exploration of preparation designs	31
4.6	Optimizing performance within a preparation design	34
5	Conclusions	35

Chapter 1

Introduction

1.1 Motivation

A *quantum state* is a collection of all possible outcomes that might result from a *measurement* of this state, with an assigned label of the probability of that outcome being observed from a measurement as opposed to the rest. The quantum state is analogous to a *noisy coin toss* [1] whose outcome is not known but we are told that it will occur with a specific probability (e.g. heads: 0.4, tails: 0.6). The quantum state will be a prediction of how many future coins will be heads or tails while a classical state, on the other hand, is deterministic i.e. either heads or tails. Modern quantum mechanical theory, thus, is inherently probabilistic. This is ordained by Bell's theorems [2–4] which postulate that we relinquish either *locality* (sometimes called causality) or *realism* from the classical macroscopic picture of the physical world when we shrink it down to the quantum scale.

Working with von Neumann's formalism [5], we as physicists have currently chosen to abandon locality: turning quantum theory into a probabilistic theory. What I mean by abandoning locality is essentially the statement: observational collapse of a quantum state from many possibilities into just one is only a consequence of measurement (i.e. *non-locality* in evolution) and at every other time we assume ignorance of the dynamics of the quantum state. The only possible way to acquire knowledge of the entire quantum state is, therefore, to work with *ensembles* i.e. multiple copies of the quantum state. The task of reconstructing a quantum state from its measurement data is called *quantum state tomography* (QST). Thus far, the quantum framework is the most accurate prescription of the microphysical world and has birthed the field of quantum information that seeks to exploit the probabilistic nature of the quantum state to store and manipulate information exponentially more efficiently than what current classical computer science promises. But the field also brings with it its own unique set of challenges. One such major challenge, characterization of quantum dynamics or quantum process tomography, will be the subject of this thesis.

Quantum process tomography (QPT) [6] is the name given to a class of protocols that reconstruct physical operations and time-evolutions of the quantum state using statistical estimation methods. Any such practicable protocol is indispensable for efficient fault-tolerant functioning of near-term quantum information processors as these will regularly require, for example, quantum error correction and validation

of quantum algorithms - both of which use process estimation schemes at their heart [6]. Moreover, QPT has proved useful in characterizing outcomes in various experiments in optical homodyne settings [7, 8].

QPT is a well-studied, exponentially hard problem in the number of quantum states; and ironically, this is also the scaling in computational power quantum processors promise for various NP-hard problems like travelling salesman-esque optimization [9] and molecular simulations [10]. It is closely linked to QST by the Choi-Jamiałkowski isomorphism theorem [11, 12], that establishes the equivalence of QPT on a system of n quantum states with QST on a system with $2n$ quantum states. QST, in turn, has been shown to be asymptotically equivalent to noisy matrix estimation via the Le Cam paradigm, a theoretical framework for statistical experiments [13]. The first correspondence allows QST estimation methods (Bayesian or not) to be applied to QPT problems without much modification. This extends the class of solutions to QPT. The second correspondence further extends the class of QST estimation procedures by allowing techniques from classical statistics to be directly applicable to QST.

All QPT protocols can be classified into two broad categories [14]:

1. *Indirect*: Indirect protocols require QST and involve reconstruction of a *transformed* initial set of input probe states by the process we wish to estimate.
2. *Direct*: Direct protocols circumvent the QST step but still require a special set of probe systems to be transformed under the action of a quantum process but whose dynamics are now characterized through quantum error correction methods (stabilizer and orbit group measurements).

Within these, the schemes can be either *online* or *offline* depending on whether the scheme can be optimized during execution. No online scheme for direct methods exists at the time of writing of this thesis. It is worth noting that indirect methods not only involve all the issues associated with QST but also more higher-level ones concomitant to QPT. However, both protocols require a *preparation* stage design involving probe input states and a *measurement* stage design involving probe measurement states or operators. The downside of direct methods is that usually the measurement and/or input designs are costly to prepare which might not be practical in experimental settings [15].

There are few direct methods in literature owing to the difficulty in designing them. Firstly, direct characterization of quantum dynamics (DCQD) proposed in ref. [16] uses certain entangled states as inputs and performs simple error-detecting measurements on the extended system making use of the extended-QST/QPT isomorphism and subsequently exploits symmetry in the extended system. Secondly, selective and efficient *incomplete* QPT proposed by Bendersky et. al. [17] exploits the measure-theoretic result that spherical t -designs allow integrals of polynomials on a sphere to be equivalent to summing over a finite set of points on the sphere [18]. Now, by setting $t = 2$, the number of indices that parametrize the quantum process, mutually unbiased (totally non-orthogonal) measurements, a natural 2-design, are performed on the extended, transformed initial state set that directly reveals information about different aspects of the state without any tomography procedure.

Both schemes are offline. DCQD still grows exponentially in resource requirement with system size but half as fast as the standard method whilst the selective efficient method, although incomplete, grows polynomially. Note that both methods essentially optimize the measurement stage of the QPT protocol and then, working backwards, design their respective input stages.

On the other hand, there is a larger variety of indirect methods in literature that just implement multiple QSTs in increasingly clever ways on extended systems. Standard QPT [19] that involves preparing a fixed set of maximally coherent and computational states and performing QST belongs to this category. All of the these methods focus on varying the measurement stage degree of freedom of QPT, constraining the input stage to allow for this variation. Some important variants include:

- Self-guided QPT (SQQPT) with the experiment determining its own state (online) [20]
- QPT with QST done through various likelihood extremization methods. These include approximate QPT that prioritizes the largest contributions in parameters that characterize a process called quantum principle component analysis [21].
- Optical continuous variable QPT following the eponymous dual prescription of quantum optics [22].
- Compressed sensing methods that aim to diagnose multiple parameters simultaneously using many-body probe and measurement designs [23].

Since almost all of these aforementioned methods focus on optimizing *measurement* design with little work done on optimizing *preparation* design has been done apart from analyses on the effects of initial environment correlations (noise) during preparation on QPT [24]. Moreover, even though many schemes propose traits that ensure robustness to noise in experimental settings like SQQPT, little work has been done to incorporate realistic noise models in QPT schema.

1.2 Thesis overview

Given the lack of literature, one naturally begs the following questions:

1. What is the precise effect of input stage variation on the final QPT scheme?
2. What quantifiable role does input design variation play in the most simplest QPT protocol (SQPT)? And can we optimize it?
3. How can we incorporate noise in QPT to account for systematic and non-quantum errors associated with QPT for a more realistic evaluation of their performances?

This thesis aims to answer points 2 and 3 directly for a single qubit and provide an outlook on how one can go about answering 1. The role of preparation design variation has not been explored beyond regularization effects (how to ensure

physicality). Here, I build a workable n -state SQPT protocol with fixed Pauli basis measurement design with decohering noise and allow the preparation design to be varied. A study on the performance of the protocol using 3 distance measures (ℓ_1 -norm, entanglement fidelity and diamond norm) during estimation reveals the diamond norm, a worse case *minimax* metric, to be ideal for comparing the true and estimated process accuracy in SQPT. Using ensembles of different sizes for various input design variations, power law fits (following the offline scaling proposed in ref. [1]) for the scaling error with ensemble sizes are obtained, with the fit coefficient serving as a proxy for the speed of convergence, of these input designs. This allows quantitatively benchmarking of various input state designs to compare their performances. An optimal input state design for the single qubit is found to be the tetrahedral state design for both noisy and non-noisy cases outperforming the default standard QPT parametrization. Necessary conditions for this optimality are symmetry and minimal size of the design and this is confirmed computationally.

This thesis is structured as follows. [Chapter 2](#) provides a theoretical background on QPT, channel representations with examples, error metrics and convex optimization. [Chapter 3](#) includes details on the SQPT computational and noise models. In [Chapter 4](#), optimal and non optimal results from varying preparation design are discussed and [Chapter 5](#) provides my holistic conclusions on the entire project and highlights promising future research avenues that were not tackled in this project due to a lack of time.

Chapter 2

Background

This chapter is a review of important physical, mathematical and computational techniques used in subsequent chapters. It presents a more linear algebraically rigorous version of the quantum state, representations of physical processes that the state can undergo i.e. *quantum channels*, and good distance measures that can allow channels to be distinguished. I have mainly followed refs. [6, 25, 26] for quantum mechanics. For ideas from convex optimization, onwards [Section 2.6](#), refs. [27, 28] was used. [Section 2.7](#) incorporates a result from the thesis [1] that will become very relevant with a QPT performance metric construction later on in [Chapter 3](#). Please refer to these for further reading.

2.1 Quantum state representation, postulates and ensembles

The quantum model is a requirement of nature for its description demanding that the *quantum* state set be bigger than classical. By definition, the quantum state $|\psi\rangle$ is a *vector* that belongs to some complex Euclidean or Hilbert¹ space $\mathcal{H} = \mathbb{C}^\Gamma$ for some alphabet Γ e.g. $\{0, 1\}$ for the quantum bit or 1-qubit. The simplest case is,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad \langle\psi| = \alpha^*\langle 0| + \beta^*\langle 1| = |\psi\rangle^\dagger \quad \forall \alpha, \beta \in \mathbb{C} \quad (2.1)$$

where a basis for the quantum state space has been defined using,

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.2)$$

The space is “dual-ed” by defining a conjugate space ² which inhabits complex conjugated and transposed reflections of the quantum state $\langle\psi| := |\psi\rangle^{*T} \equiv |\psi\rangle^\dagger$. This allows a natural norm or complex scalar map from the duals to vectors to be their inner product. It is normalized to 1 by requirements of probability,

$$\langle\psi|\psi\rangle = \alpha^*\alpha\langle 0|0\rangle + \beta^*\beta\langle 1|1\rangle + \alpha^*\beta\langle 0|1\rangle + \beta^*\alpha\langle 1|0\rangle = |\alpha|^2 + |\beta|^2 = 1 \quad (2.3)$$

¹to allow continuity

²always exists by the Riez-representation theorem

2 Background

where $\langle 0|1\rangle = \langle 1|0\rangle = 0$ by orthogonality of the linearly independent basis vectors as $\text{span}(|\psi\rangle) = \{|0\rangle, |1\rangle\}$. Another useful way to write (2.1) is,

$$|\psi\rangle = |\psi(\theta, \phi)\rangle = e^{i\gamma} \left(\cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \right) \quad (2.4)$$

where $\gamma, \theta, \phi \in \mathbb{R}$ and $e^{i\gamma}$ is a global phase factor and can usually be ruled out after some consideration. For 2-qubits, we have $\text{span}(|\psi\rangle_{AB}) = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ where $|00\rangle \equiv |0\rangle_A \otimes |1\rangle_B$. Qubit A and qubit B now inhabit the bigger $\mathbb{C}^{\Gamma \times \Gamma} \equiv \mathbb{C}_A^2 \otimes \mathbb{C}_B^2$. We can consider them as subsystems of the 2-qubit. Both vector spaces are independent. An n-qubit inhabits $\mathbb{C}_1^2 \otimes \mathbb{C}_2^2 \cdots \otimes \mathbb{C}_n^2 \equiv \mathbb{C}^{2^{\otimes n}}$. Generally, for linear operators (including the vector) \hat{A} and \hat{B} , using hats to denote these, the tensor product is defined as,

$$\hat{A} \otimes \hat{B} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} \otimes \hat{B} = \begin{pmatrix} a_{11}\hat{B} & a_{12}\hat{B} & \dots & a_{1n}\hat{B} \\ a_{21}\hat{B} & a_{22}\hat{B} & \dots & a_{2n}\hat{B} \\ \vdots & \vdots & \ddots & \\ a_{m1}\hat{B} & \dots & \dots & a_{mn}\hat{B} \end{pmatrix} \quad (2.5)$$

The above is not a general representation of the quantum state. We also need to consider a *convex* combination of *pure* states like (2.1), something that can be written as $(1 - \lambda)\hat{u} + \lambda\hat{v}$ for some $0 \leq \lambda \leq 1 \in \mathbb{R}$ where \hat{u}, \hat{v} are pure states. This is because we wish to consider quantum states that could be combinations of many such pure states e.g. possibly as a single system pure state in the presence of another larger system like the environment or something smaller like another physical system. The stronger definition is a *density operator* $\hat{\rho} \in D(\mathcal{H})$ [5] where $D(\mathcal{H})$ is a space of positive semi-definite operators with trace $\text{Tr } \hat{\rho} = 1$. Positive semi-definite operators $P(\mathcal{J})$ are the space of all linear operators \hat{B} that can be written like $\hat{B}^\dagger \hat{B}$. And the trace 1 requirement means $\text{Tr } \hat{B} := \sum_i \hat{B}_{ii} = 1$. The former condition ensures that the eigenvalue spectrum is real and positive i.e measurable and the latter ensures the spectrum sums to 1 conserving probability, what we're modelling. The general quantum state is given by,

$$\hat{\rho} = \sum_{a \in \Gamma} p(a) |\psi\rangle_a \langle \psi|_a = \sum_{a \in \Gamma} p(a) \hat{\rho}_a \quad (2.6)$$

where $p(a)$ is the probability associated with each pure state $|\psi\rangle_a$ which can be written as a linear operator as $\hat{\rho}_a := |\psi\rangle_a \langle \psi|_a$ e.g. $|0\rangle\langle 0| = (1, 0) \cdot (1, 0)^T$ is written as an outer product. All linear operators have a convenient matrix representation and the tensor product extension to include more states can be applied identically to $\hat{\rho}$. The qubit (2.1) can be re-written now,

$$\hat{\rho} = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix} \quad \text{where } \text{Tr } \hat{\rho} = |\alpha|^2 + |\beta|^2 = 1 \quad \text{and } \hat{\rho}^\dagger = \hat{\rho} \quad (2.7)$$

The diagonal elements are accessible classical states that the qubit can occupy when we observe it. These are equivalent to a probabilistic classical state set. One cannot actually construct the off-diagonal elements called coherences on their own as the

2 Background

above two constraints forbid it. It is, however, possible to construct physical states in other *bases* and use linear combinations of those to reconstruct the off-diagonal element e.g.

$$|0\rangle\langle 1| = |+\rangle\langle +|_x + |-\rangle\langle -|_y - \frac{1+i}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \text{ where } |\pm\rangle_x = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}, \quad |\pm\rangle_y = \frac{|0\rangle \pm i|1\rangle}{\sqrt{2}}$$

Pure states are extreme points of the convex set containing all $\hat{\rho}$ where $\lambda = 0, 1$. Everything else is dubbed *mixed*³.

The pure density operator can also be decomposed using a 3 real variable parametrization (r_x, r_y, r_z) and basis matrices⁴, as, (this is analogous to (2.4))

$$\hat{\rho}(\vec{r}) := \frac{1}{2} \left(\hat{1} + \vec{r} \cdot \vec{\sigma} \right) = \frac{1}{2} \left(\hat{1} + r_x \hat{\sigma}_x + r_y \hat{\sigma}_y + r_z \hat{\sigma}_z \right) \quad (2.8)$$

where

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \hat{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2.9)$$

These operators have rank 2 (the dimension of their image is 2 or more simply the number of linearly independent columns inside them) with eigenvalues ± 1 . These can be written in terms of rank 1 pure states (their eigenstates) as,

$$\sigma_x = +1|+\rangle\langle +|_x - 1|-\rangle\langle -|_x \quad \sigma_y = +1|+\rangle\langle +|_y - 1|-\rangle\langle -|_y \quad \sigma_z = +1|0\rangle\langle 0|_z - 1|1\rangle\langle 1|_z \quad (2.10)$$

It is only convention to set the z-eigenstates as the $\{|0\rangle, |1\rangle\}$ computational basis. Both (2.4) and (2.8) allow us to represent the pure state on the surface of a (2-)sphere called the Bloch sphere which is a very useful geometric picture.

The postulates that govern the dynamics of the qubit can now be stated,

1. Any isolated quantum system is completely described by its density operator $\hat{\rho}$ (2.6).
2. The isolated quantum system evolves from say time t_1 to t_2 under *unitary* transformations \hat{U} that are obtained from solving the Schrödinger equation,

$$\hat{\rho} \rightarrow \hat{U}\hat{\rho}\hat{U}^\dagger \quad \text{from solving} \quad i\frac{\partial\hat{\rho}}{\partial t} = \hat{H}\hat{\rho} - \hat{\rho}\hat{H}$$

where $\hat{U} = U(t_2 - t_1)$ and $\hat{U}(0) = \hat{1}$ and $\hat{U}^\dagger\hat{U} = \hat{1}$ to ensure reversibility [5]. For an isolated free state $\hat{U} = e^{-i\hat{H}t}$ for some Hamiltonian \hat{H} corresponding to rotations of the state about an axis defined by the \hat{H} .

3. Quantum measurements are described by a set of positive definite measurement operators⁵ $\{\hat{P}_m\}$ that usually also satisfy the condition $\hat{P}^2 = \hat{P} = \hat{P}^\dagger$ like the Pauli operator eigenstates above. The state after observation is “collapsed” with probability $p(m|\rho)$,

$$\hat{\rho} \xrightarrow{\hat{P}_m} \frac{\hat{P}_m^\dagger \hat{\rho} \hat{P}_m}{\text{Tr } \hat{P}_m^\dagger \hat{P}_m \hat{\rho}} \quad \text{where} \quad \sum_m \hat{P}_m^\dagger \hat{P}_m = \hat{1} \quad \text{and} \quad p(m|\rho) = \text{Tr } \hat{P}_m^\dagger \hat{P}_m \hat{\rho} \quad (2.11)$$

³The maximally mixed state is normalized identity $\hat{1}/2$ which is at the center of the SU(2) Lie group to which $\hat{\rho}$ belongs.

⁴SU(2) basis generators

⁵a result of Naimark’s theorem from functional analysis

4. State sets of compound systems (recall earlier example) is the tensor product of state spaces of the component physical systems, which can be treated and interacted with individually.

Postulate 3 necessitates that all the information of a quantum state's elements cannot be acquired through a single measurement. We need to consider measurements of multiple copies of the state i.e. an ensemble, to reconstruct a probability distribution that parametrizes the individual elements $\hat{\rho}_{ij}$. Postulate 4 says one can make partial measurements on composite systems and there can be partial knowledge of the subsystem but not the whole through the use of a partial trace. Suppose one applies a linear operator \hat{T} on a subsystem of a composite system $K = K_a \otimes K_b$. The effect of measurement is then, as $\text{Tr} \equiv \text{Tr}_a \text{Tr}_b$

$$\text{Tr} [(\hat{T}_a \otimes \hat{1}_b) \hat{K}] = \text{Tr}_a [\hat{T}_a \text{Tr}_b \hat{K}] \quad (2.12)$$

due to nothing but the linearity of the trace and the fact that both subspaces A and B are interacted with separately e.g. through product measurements $\hat{\sigma}_x \otimes \hat{\sigma}_y$ where $\hat{1}$ reveals nothing. The partial trace completely removes the traced over vector space after application and projects us into a subspace.

Almost all of quantum operators including $\hat{U}, \hat{P}, \hat{\rho}$ are by construction normal $\hat{N}^\dagger \hat{N} = \hat{N} \hat{N}^\dagger$ – a subclass of which are Hermitian operators like $\hat{\rho} = \hat{\rho}^\dagger$ – which allows them to have a diagonal *spectral* decomposition like (2.6) (e.g. see ref. [6] for proof). Also from postulate 4, any mixed state turns into a pure state by extending it through tensor product composition with another auxiliary system. After partial tracing, this gives us back the mixed state.

2.2 Channel representations

A *quantum channel* \mathcal{E} is a representation of discrete changes that can be physically realized on a quantum state, examples include the measurement and the unitary transform. So far we have considered the evolution of a state in isolation i.e. in the case of a closed quantum system. But we now have enough machinery at our disposal talk about open quantum systems using the quantum channel. Mathematically, the channel is a linear map from any square linear operator to another,

$$\mathcal{E} : L(\mathcal{X}) \rightarrow L(\mathcal{Y}) \quad \text{and transforms} \quad \hat{\rho}' = \hat{\rho} \xrightarrow{\mathcal{E}} \mathcal{E}(\hat{\rho})$$

that must satisfy two properties,

1. Its action must be *completely positive*: Not only do we need $\mathcal{E}(\hat{\rho})$ for some $\hat{\rho}$ in \mathcal{H}_{in} to be positive but also $(\mathcal{E} \otimes \hat{1}_a) \circ \hat{\rho}$ for some $\hat{\rho}$ in $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_a$.
2. Its action must be *trace preserving* or (slightly weaker) at least *trace non-increasing*. $0 \leq \text{tr } \mathcal{E}(\rho) \leq 1$ i.e. the operation has a finite probability of occurrence.

2 Background

In condition 1, the latter does not imply the former. A contradictory example comes from the positive partial transposition map on a 2-qubit state $\hat{\rho} \otimes \hat{\sigma} \rightarrow \hat{\rho} \otimes \hat{\sigma}^T$. On the state $|b\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ as a density operator, this is

$$\hat{\rho} = |b\rangle\langle b| = \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}} \right) \left(\frac{\langle 00| + \langle 11|}{\sqrt{2}} \right) = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{1 \otimes T} \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

which obtains a negative eigenvalue $-\frac{1}{2}$.

The map must be linear such that $\mathcal{E}\left(\sum_i p_i |\psi_i\rangle\langle\psi_i|\right) = \sum_i p_i \mathcal{E}(|\psi_i\rangle\langle\psi_i|)$ so that the operation ensures that the transformation of a mixed state is the same as the probabilistic sum of the transformations of component pure states.

These constraints preserve the physical structure of density operators. Consider an equivalent, more physical interpretation. Assume that the system $\hat{\rho}$ can be represented as composite tensor product state with a pure environment state $\hat{\rho}_{\text{env}} = |e_0\rangle\langle e_0|$. The environment can be made as large as we wish until it becomes pure (using postulate 4), so there is no loss of generality. Denote $|e_k\rangle$ as the orthonormal basis for $\hat{\rho}_{\text{env}}$ for the environment system. Suppose that combined, the system and the environment evolve via a closed unitary evolution dictated by postulate 2. Assume that only the environment can interact with the state and we assume that interaction is equivalent to making measurements via postulate 3. After a certain time, various elements of system $\hat{\rho}_k$ will be given by,

$$\mathcal{E}(\hat{\rho})_k \propto \text{Tr}_{\text{env}} \left[|e_k\rangle\langle e_k|_{\text{env}} \hat{U} (\hat{\rho} \otimes \hat{\rho}_{\text{env}}) \hat{U}^\dagger |e_k\rangle\langle e_k|_{\text{env}} \right] = \langle e_k |_{\text{env}} \hat{U} (\hat{\rho} \otimes \hat{\rho}_{\text{env}}) \hat{U}^\dagger |e_k\rangle_{\text{env}} \quad (2.13)$$

where the Tr_{env} is the partial trace over environment and after abstracting away

$$\hat{E}_k = \langle e_k | \hat{U} | e_0 \rangle \quad \hat{E}_k^\dagger = \langle e_0 | \hat{U}^\dagger | e_k \rangle \quad (2.14)$$

as a Krauss operator and its conjugate-transpose. Normalizing it gives,

$$\mathcal{E}(\hat{\rho})_k = \frac{\hat{E}_k \hat{\rho} \hat{E}_k^\dagger}{\text{Tr } \hat{E}_k \hat{\rho} \hat{E}_k^\dagger} \quad (2.15)$$

so the channel can be given a normal operator sum representation by considering all environment-aware unitary state evolutions as,

$$\mathcal{E}(\hat{\rho}) = \sum_k \mathcal{E}(\hat{\rho})_k = \sum_k \left(\text{Tr } \hat{E}_k \hat{\rho} \hat{E}_k^\dagger \right) \hat{\rho}_k = \sum_k p(k) \hat{\rho}_k = \sum_k \hat{E}_k \hat{\rho} \hat{E}_k^\dagger \quad (2.16)$$

Figure 2.1 visually contrasts quantum operations on open and closed systems. Note that by taking the trace of (2.16), $\text{tr}(\mathcal{E}(\hat{\rho})) \leq 1 \implies \sum_k \hat{E}_k^\dagger \hat{E}_k \leq \mathbb{I}$. When $\sum_k \hat{E}_k^\dagger \hat{E}_k = \mathbb{I}$, the map is completely positive and trace preserving (CPTP) and when $k = 1$ one arrives at unitary dynamics of a closed system.

One fatal weakness of this formalism (2.16) is that it assumes there are no correlations between the system and environment before evolution starts occurring. If

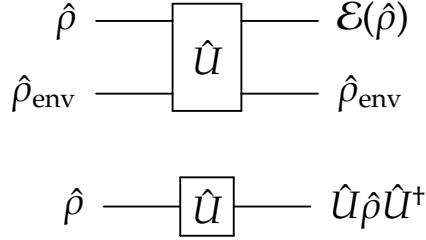


Figure 2.1: Quantum circuits contrasting the action of a unitary (quantum gate) on open (top) and closed (bottom) systems ρ

there are, then the evolution is not CP [24]. This is a signature of *entanglement*, non-classical correlations between states in composite systems. Classical states are not composed from superpositions of possible states, at the individual system level, so non-zero off-diagonals in composite systems like $|b\rangle\langle b|$ lead to non-classical physical results dubbed entanglement. Alternative ways to deal with this use quantum discord and assignment mapping formalism [29, 30]. We will not concern ourselves with this technicality in this thesis.

Another fact to bear in mind is that the Krauss operators $\{\hat{E}_k\}$ are not unique and there are many unitarily transformed operators for the same dynamics. We will attempt to remedy this by introducing a major correspondence between quantum states and channels called the *Choi-Jamiałkowski isomorphism*. Consider the state,

$$|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_i^{d-1} |i\rangle \otimes |i\rangle$$

By using the completely positive property of the channel \mathcal{E} , we can define a new state,

$$\hat{\rho}_{\mathcal{E}} = (\mathcal{E} \otimes \mathbb{1}) |\Phi^+\rangle \langle \Phi^+| \leftrightarrow (\mathbb{1} \otimes \mathcal{E}) |\Phi^+\rangle \langle \Phi^+| \quad (2.17)$$

Thus, characterizing the dynamics of an arbitrary 1-qubit is equivalent to characterizing the dynamics of just one half of the state $|\Phi^+\rangle$. This is a very important result as the channel-state duality is exploited in various areas of quantum information.

The final representation that will be useful is the so-called *Superoperator* representation of the channel. By vectorizing the state $\hat{\rho}$,

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \longrightarrow \vec{\rho} \leftrightarrow |\rho\rangle = (\rho_{00}, \rho_{01}, \rho_{10}, \rho_{11})^T \quad \text{where} \quad \rho_{ij} = \text{Tr } \hat{h}_i \hat{\rho}$$

and noting that $\{\hat{h}_i\}$ forms an orthonormal basis ⁶ that is not Pauli. Then, \mathcal{E} can be converted into a super-operator that satisfies,

$$|\mathcal{E}(\rho)\rangle = \bar{\mathcal{E}} |\rho\rangle \quad (2.18)$$

where

$$\bar{\mathcal{E}} = \mathcal{E}_{ij} |i\rangle\langle j| \quad \text{and} \quad \mathcal{E}_{ij} = \text{Tr } \hat{h}_i^\dagger \mathcal{E}(\hat{h}_j) \quad \text{and} \quad \mathcal{E}(\hat{k}) = \sum_{ij} \hat{h}_i \mathcal{E}_{ij} \hat{h}_j$$

⁶for SU(2)

2 Background

and now the columns of the superoperator $\bar{\mathcal{E}}$ essentially encode the entries $\mathcal{E}(\hat{\rho})$. By knowing Roth's lemma for linear operators,

$$\hat{A}\hat{B}\hat{C} \iff \hat{A} \otimes \hat{C}^T |B\rangle \quad (2.19)$$

we can arrive at a vectorized version of the operator-sum representation (2.16),

$$|\mathcal{E}(\hat{\rho})\rangle = \sum_k E_k \otimes E_k^* |\rho\rangle \quad (2.20)$$

and of the Choi state $\hat{\rho}_{\mathcal{E}}$ in (2.17) by partial transposing $\bar{\mathcal{E}}$ using (2.19),

$$(\hat{\rho}_{\mathcal{E}})_{ab,\mu\nu} = (\bar{\mathcal{E}})_{ab,\mu\nu}^{pT} = (\bar{\mathcal{E}})_{a\mu,b\nu} = \begin{pmatrix} \bar{\mathcal{E}}_{00} & \bar{\mathcal{E}}_{01} & \bar{\mathcal{E}}_{10} & \bar{\mathcal{E}}_{11} \\ \bar{\mathcal{E}}_{02} & \bar{\mathcal{E}}_{03} & \bar{\mathcal{E}}_{12} & \bar{\mathcal{E}}_{13} \\ \bar{\mathcal{E}}_{20} & \bar{\mathcal{E}}_{21} & \bar{\mathcal{E}}_{30} & \bar{\mathcal{E}}_{31} \\ \bar{\mathcal{E}}_{22} & \bar{\mathcal{E}}_{23} & \bar{\mathcal{E}}_{32} & \bar{\mathcal{E}}_{33} \end{pmatrix} \quad (2.21)$$

This superoperator representation will be very useful for comparing and evaluating actions of quantum channels. Also, complete positivity of the eigenvalue spectrum of (2.21) will provide one reasonable test for ensuring that the channel is CPTP.

2.3 Major quantum operations in the geometric picture

Following (2.8), we can represent most quantum operations or channels as affine transformations of the Bloch vector \vec{r} ,

$$\vec{r} \xrightarrow{\mathcal{E}} \hat{M}\vec{r} + \vec{c} \quad (2.22)$$

This can be seen immediately by parametrizing expanding the Krauss operators \hat{E}_k in terms of the Pauli operators,

$$\hat{E}_k = \alpha_i \mathbb{I} + \sum_{j=1}^3 a_{ij} \hat{\sigma}_j \quad (2.23)$$

then inserting (2.23) and $\hat{\rho}(\vec{r})$ (2.8) into the operator-sum \mathcal{E} (2.16) and expanding yields the form of elements of \hat{M}_{jl}

$$\hat{M}_{jl} = \sum_q \left(a_{qj} a_{ql}^* + a_{qj}^* a_{ql} + \left(|\alpha_a|^2 - \sum_t a_{qt} a_{qt}^* \right) \delta_{jl} + i \sum_t \epsilon_{jlt} \left(\alpha_q a_{qt}^* - \alpha_q^* a_{qt} \right) \right) \quad (2.24)$$

where $\epsilon_{ijk} = 1$ if $ijk = 123, 312, 231$ and $\epsilon_{ijk} = -1$ if $ijk = 132, 321, 213$ and zero otherwise, is the fully antisymmetric tensor. And the kronecker delta $\delta_{ij} = 1$ if $i = j$ and zero otherwise. Likewise the components for the translation vector c_l are,

$$c_l = 2i \sum_t \sum_{jp} \epsilon_{jpl} a_{tj} a_{tp}^*$$

The forms of \hat{M} in (2.24) and \vec{c} tell us that both are real. We can decompose $\hat{M} = \hat{U}|\hat{M}|$ where \hat{U} is a unitary and $|\hat{M}|$ is real Hermitian matrix ($|\hat{M}| = \sqrt{\hat{M}^\dagger \hat{M}}$) in its polar

2 Background

form. Since \hat{M} is real, \hat{U} can also be taken to be real in which case it will just be an orthogonal matrix satisfying $\hat{U}^T \hat{U} = \mathbb{1}$. Its determinant can then be taken to be 1. Composing both operations yielding implies that it is simply a rotation operation governed by \hat{U} with deformations along each axis given by the diagonal $|\hat{M}|$ (using spectral decomposition). All channels can't be represented in this framework like those that have reflection or inversion effects at points or "collapses" associated with measurements on \vec{r} e.g. like the CNOT gate given by,

$$\hat{U}_{\text{CNOT}} = |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 11| + |11\rangle\langle 10| = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \hat{\sigma}_x$$

that flips the computational state in the second system if the state in the first system is $|1\rangle$.

Coming to the channels that can be represented, the first that I will present, is the symmetric *depolarizing* channel that is a representation of a constant noise addition to a quantum system,

$$\mathcal{P}(p, \hat{\rho}) = \mathcal{P}(p) = p \frac{\hat{\mathbb{1}}}{2} + (1-p) \hat{\rho} = \left(1 - \frac{3}{4}\right) \hat{\rho} + \frac{p}{4} \left(\hat{\sigma}_x \hat{\rho} \hat{\sigma}_x + \hat{\sigma}_y \hat{\rho} \hat{\sigma}_y + \hat{\sigma}_z \hat{\rho} \hat{\sigma}_z\right) \quad (2.25)$$

where the last equality follows from the identity,

$$\frac{\mathbb{1}}{2} = \frac{\hat{\rho} + \hat{\sigma}_x \hat{\rho} \hat{\sigma}_x + \hat{\sigma}_y \hat{\rho} \hat{\sigma}_y + \hat{\sigma}_z \hat{\rho} \hat{\sigma}_z}{4}$$

and dependence on the state will now be implicit. The Krauss operator set $\{\hat{E}_k\}$ is $\{\sqrt{1-3p/4}\hat{\mathbb{1}}, \sqrt{p}\hat{\sigma}_x/2, \sqrt{p}\hat{\sigma}_y/2, \sqrt{p}\hat{\sigma}_z/2\}$. Hence, the Bloch vector is symmetrically shrunk by the application of $\mathcal{P}(p)$ tending towards the invariant state at the center which is $\hat{\mathbb{1}}/2$. Another channel is the *dephasing* channel $\mathcal{H}(p)$ that corresponds to the loss of coherence elements (off-diagonals),

$$\mathcal{H}(p) = p |0\rangle\langle 0| \hat{\rho} |0\rangle\langle 0| + (1-p) |1\rangle\langle 1| \hat{\rho} |1\rangle\langle 1| \quad (2.26)$$

with Krauss operators $\{\sqrt{p}\hat{\mathbb{1}}, \sqrt{1-p}\hat{\sigma}_z\}$. Information about the x and y axes is getting lost and corresponds to shrinking of \vec{r} only for these components. This corresponds to the mapping $\vec{r} \rightarrow (r_x(1-2p), r_y(1-2p), r_z)$ and the z axis states are invariant. Analogous dephasing operations along other axes can also be written in a similar way. The final channel that I will explicitly state is the general amplitude damping channel $\mathcal{D}(\gamma)_G$. It is given by

$$\mathcal{D}(\gamma)_G = \hat{E}_0 \hat{\rho} \hat{E}_0 + \hat{E}_1 \hat{\rho} \hat{E}_1 + (\hat{E}_2 \hat{\rho} \hat{E}_2^\dagger + \hat{E}_3 \hat{\rho} \hat{E}_3^\dagger) \quad (2.27)$$

where

$$\hat{E}_0 = \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \quad \hat{E}_1 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \quad \hat{E}_2 = \sqrt{1-p} \begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix} \quad \hat{E}_3 = \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix}$$

and the Krauss operators coincide with the operators defined above. This represents dissipation to an environment and tends towards the invariant state,

$$\hat{\rho}_{\text{inv}} = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}$$

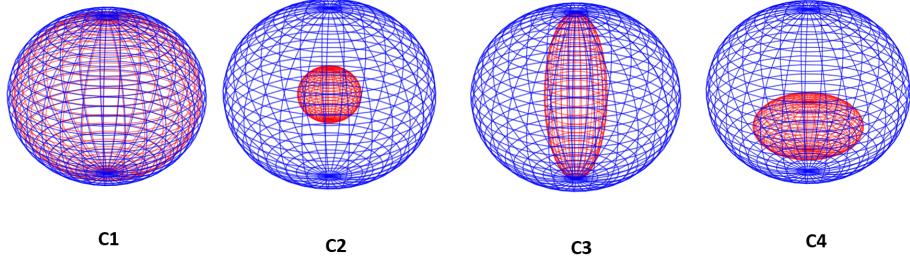


Figure 2.2: Example single qubit channels are shown. C1: a unitary rotation; C2: depolarizing $\mathcal{P}(0.3)$; C3: dephasing $\mathcal{H}(0.4)$; C4: amplitude damping $\mathcal{D}(0.3)$

Its general effect is $\vec{r} \rightarrow (r_x \sqrt{1-\gamma}, r_y \sqrt{1-\gamma}, \gamma(2p-1) + r_z(1-\gamma))$. We can denote a simpler more specific version with $p = 1$ by $\mathcal{D}(\gamma)$ that corresponds to physical phenomena like spontaneous emission of the excited state in a two level atom or asymptotic decay of some excited state to the ground state permanently. All the above maps are trace preserving. And in fact, all CPTP maps tend to an invariant state⁷. Figure 2.2 displays the effects of some channels on the Bloch sphere.

2.4 Quantum state and process/channel tomography

Suppose you were given M copies of an unknown n -qubit state $\hat{\rho}$ and were asked what it was. You would perform what is called *quantum state tomography* (QST). For an n -qubit system, this can be performed by expressing ρ in terms of the Pauli basis using postulates 3 and 4,

$$\rho = \sum_{v_1, v_2, \dots, v_n}^3 \frac{\text{tr}(\sigma_{v_1} \otimes \dots \otimes \sigma_{v_n} \rho) (\sigma_{v_1} \otimes \dots \otimes \sigma_{v_n})}{2^n} \quad (2.28)$$

where each $v_i = 0, 1, 2, 3$ indicate \mathbb{I}, X, Y, Z . In principle, we can estimate $\text{tr}(\sigma_{v_1} \otimes \dots \otimes \sigma_{v_n} \rho)$ by performing the equivalent measurement on m copies of ρ . Since product measurements can be treated independently, the measurement probability distributions can be treated as independent random variables. The central limit theorem then applies and the error or deviation in the measurements from truth is gaussian and decreases as $1/\sqrt{M}$.

By formalizing what we mean by a quantum channel (a CPTP dynamical map), we can now understand methods of characterizing such a channel. Note that a channel can represent just a single gate operation like in figure 2.1 or indeed a whole circuit with multiple channels applying on a state. It should be clear now that by characterize, we mean estimating the Krauss operator set $\{E_k\}$.

Quantum process tomography (QPT) is an isometric extension of quantum state tomography. The easiest way to do it would be to use the state-channel correspondence (2.17) and perform QST on the choi state $\hat{\rho}_{\mathcal{E}}$. This is an example of ancilla assisted process tomography (AAPT) by adding an artificial extension to a probe

⁷a consequence of Schauder's fixed point theorem for convex, bounded and continuous linear operators

physical system, we can apply the operation we wish to diagnose on one half of the extended system, the choi state $\hat{\rho}_{\mathcal{E}}$, and just study the changes on just this one state through QST. This then results in a set of linear equations that constrain various elements of $\mathcal{E}(\hat{\rho})$. It was first demonstrated for electromagnetic fields in an experimental twin beam homodyne setup [31]. However, qubit-preparation wise, this is expensive and often impossible to scale up.

Another way is to encode all the possible transformations on the elements of $\hat{\rho}$ using the standard quantum process tomography (SQPT) method. Experimentally, this can be done in a fixed measurement operator basis $\{\hat{E}_i\}$ where a general Krauss operator element \hat{E}_k can be decomposed (like in (2.23)),

$$\hat{E}_k = \sum_i e_{km} \hat{E}_m \quad \forall e_{km} \in \mathbb{C}$$

Equation (2.16) can be rewritten,

$$\mathcal{E}(\rho) = \sum_{mn}^{d^2-1} \tilde{E}_m \rho \tilde{E}_n^\dagger \chi_{mn} \quad (2.29)$$

where

$$\chi_{mn} = \sum_q e_{mq} e_{qn}^* \quad (2.30)$$

is the positive hermitian $\hat{\chi}$ matrix that will contain $d^4 - d^2$ independent real parameters and $d = \dim \mathcal{H}_{\text{in}} = \text{tr}(\hat{E}_m^\dagger \hat{E}_n)$ if $m = n$. Preparing a set of linearly independent input states $\{\hat{\rho}_j = |m\rangle \langle n|\}_{j=0}^{d^2-1}$ allows one to estimate $\mathcal{E}(|m\rangle \langle n|)$ (decomposed spectrally) using quantum state tomography. The following is the standard parametrization used for off-diagonal $m \neq n$ elements (similar to (2.1), using linearity of the channel,

$$\mathcal{E}(|m\rangle \langle n|) = \mathcal{E}(|+\rangle \langle +|) + i\mathcal{E}(|-\rangle \langle -|) - \frac{1+i}{2} (\mathcal{E}(|m\rangle \langle m|) + \mathcal{E}(|n\rangle \langle n|)) \quad (2.31)$$

where

$$|+\rangle \equiv \frac{|m\rangle + |n\rangle}{\sqrt{2}} \quad |-\rangle \equiv \frac{|m\rangle + i|n\rangle}{\sqrt{2}}$$

By writing $\mathcal{E}(\hat{\rho}_j) = \sum_k \lambda_{jk} \hat{\rho}_k$ and using the same trick in (2.28) we find that

$$\lambda_{jk} = \text{Tr } \hat{\rho}_j \mathcal{E}(\hat{\rho}_k)$$

for $\tilde{E}_j = \hat{\rho}_j$ as natural measurement operators and that

$$\hat{E}_m \hat{\rho}_j \hat{E}_n^\dagger = \sum_k \hat{B}_{mn,kj} \hat{\rho}_k \quad (2.32)$$

This is a fixed encoding of all elements $\hat{\rho}_{jk}$. Now substituting (2.32) into (2.29) yields

$$\sum_{mn} \hat{B}_{mn,kj} \hat{\chi}_{mn} = \hat{\lambda}_{jk} \implies \bar{\mathbf{B}} |\chi\rangle = |\lambda\rangle$$

in superoperator form following the convention from (2.18) giving $|\chi\rangle = \bar{\mathbf{B}}^{-1} |\lambda\rangle$ through a linear inversion operation that exists by the fixed construction of $\bar{\mathbf{B}}$ (2.32).

This is the standard QPT protocol where d^4 expectation values need to be computed for d^2 prepared basis input states $\{\rho_j\}$ with their corresponding d^2 fixed basis measurement operators $\{\tilde{E}_k\}$.

2.5 Channel error and distance metrics

Because of the non-uniqueness or unitary freedom of Krauss operators, devising measures of distances between channels is done by comparing the changes they imprint on probe states $\hat{\rho}$. This, in essence, is comparing distances between measurements' probability distributions. One measure is the *entanglement fidelity* [6, 32]. We start with the fidelity metric $F(\hat{\rho}, \hat{\sigma})$ for density operators which has been proven to have nice comparison-wise properties,

$$F(\hat{\rho}, \hat{\sigma}) \equiv \text{Tr} \left[\sqrt{\hat{\rho}^{\frac{1}{2}} \hat{\sigma} \hat{\rho}^{\frac{1}{2}}} \right] \equiv \max_{|\psi\rangle, |\varphi\rangle} |\langle \psi | \varphi \rangle| \quad (2.33)$$

where the final equivalence is proven by Uhlmann's theorem on a larger Hilbert space of states $\hat{\rho}, \hat{\sigma}$ as kets $|\psi\rangle, |\varphi\rangle$ and the maximum is over all possible extensions of $\hat{\rho}, \hat{\sigma}$. This is another manifestation of the state-channel duality (2.17). Note: square roots are taken on their spectrally decomposed eigenvalue form. The fidelity metric allows for unitary freedoms in the states, observe the inner product on the right and is bounded between 0 and 1 (a metric property). I will refer the reader to refs. [6, 25] for the proofs. Working in the superoperator representation of a channel \mathcal{E} , we can define the analogous entanglement fidelity for the channel using a modified version of the Choi state following [33],

$$F(\hat{\rho}, \mathcal{E}) = \min_{\hat{\rho}} F(\hat{\rho}, \mathcal{E} \otimes \hat{1}(\hat{\rho})) \quad (2.34)$$

where the minimization is over all possible extensions and dynamics for the remainder of the system to get the *worst possible value* the fidelity could have, regardless of the actual dynamics in remaining bits. The interpretation is that this metric measures disturbances in entanglement of a larger system caused by a channel transformation on a part of it. Taking the difference between the fidelities of two channels allows one to quantify a distance between them.

Another measure is the diamond norm [34] (Dnorm) or the completely bounded trace norm, which is, arguably, a better measure to distinguish channels, in the sense that it is multiplicative if we consider applying different channel maps on subsystems of composite quantum systems that are allowed to be mixed. This effectively allows its usage to be parallelized but it is essentially very closely related to the entanglement fidelity in its properties and its method of computation, which I will postpone to the next section. Without going into too much mathematical detail that is outside the scope of this thesis, the diamond norm of the CPTP maps \mathcal{E} and Φ is given by, $\|\mathcal{E}(\hat{\rho}) - \Phi(\hat{\rho})\|_\diamond$ by solving,

$$\|\mathcal{E}(\hat{\rho}) - \Phi(\hat{\rho})\|_\diamond = \max_{\text{Tr}\rho \leq 1} \text{Tr} (\mathcal{E}(\hat{\rho}) - \Phi(\hat{\rho})) \otimes \hat{1}_n \quad (2.35)$$

Another representation of the Dnorm in terms of $F(., .)$ is [35, 36],

$$\|\mathcal{E} - \Phi\|_\diamond = \max_{\hat{\rho}, \hat{\sigma}} F(\mathcal{E}(\hat{\rho}), \Phi(\hat{\sigma})) \quad (2.36)$$

where again the worse possible difference between channels is evaluated over all positive semi-definite operator space. Both $F(., .)$ and Dnorm are stabilizer norms

with the stabilizer property for the norm is defined as follows for a distance metric $d(\cdot)$,

$$d(\hat{1}_n \otimes \Phi, \hat{1}_m \otimes \psi) \leq d(\Phi, \psi) \quad \forall n, m \geq 1$$

This makes sense due to redundancy in physical maps on open quantum systems.

2.6 Semi-definite program to compute the diamond norm

A mathematical formulation of the optimization problem in (2.36) is called a *program*. This will be a semi-definite program since it pertains to the positive semi-definite operator space to which the CPTP channel belongs. Essentially, a semi-definite program is specified by the triple $(\hat{\Xi}, \hat{A}, \hat{B})$ where $\hat{\Xi} : \hat{A} \rightarrow \hat{B}$ is a Hermiticity preserving map and \hat{A} and \hat{B} are Hermitian operators and $\hat{A} \in \text{Herm}(\mathbb{C}^n)$ and $\hat{B} \in \text{Herm}(\mathbb{C}^m)$ that need not have the same dimensions n, m . The triple is associated with two optimization problems called the *primal* and *dual*,

Primal Problem

$$\begin{aligned} \text{maximize : } & \text{Tr } \hat{A}^\dagger \hat{X} \\ \text{subject to : } & \hat{\Xi}(\hat{X}) = \hat{B} \\ & \hat{X} \in \mathcal{P}(\mathbb{C}^n) \end{aligned}$$

Dual Problem

$$\begin{aligned} \text{minimize : } & \text{Tr } \hat{B}^\dagger \hat{Y} \\ \text{subject to : } & \hat{\Xi}^\dagger(\hat{Y}) \geq \hat{A} \\ & \hat{Y} \in \text{Herm}(\mathbb{C}^m) \end{aligned}$$

where if the positive definite operator (eigenvalues strictly greater than 0) \hat{X} satisfies the above equality $\hat{\Xi}(\hat{X}) = \hat{B}$ is called *primal feasible* and likewise for some *dual feasible* \hat{Y} . The objective functions in the first line are defined using the Hilbert-Schmidt inner product, which for operators \hat{A}, \hat{B} in Hilbert space, is

$$\langle A | B \rangle = \text{Tr } \hat{A}^\dagger \hat{B} \tag{2.37}$$

We also define *primal optimal* α and its dual β , solutions to the above problems, as,

$$\alpha = \sup_{\hat{X} \in \mathcal{P}} \text{Tr } \hat{A}^\dagger \hat{X} \quad \beta = \inf_{\hat{Y} \in \text{Herm}(\mathbb{C}^m)} \text{Tr } \hat{B}^\dagger \hat{Y} \tag{2.38}$$

$\alpha \leq \beta$ holds for all semi-definite programs and is called weak duality. Strong duality is defined by $\alpha = \beta$ which holds for almost all practical programs like ours [28]. In this paradigm, the diamond norm can be formulated, skipping some lengthy linear algebra from Watrous et. al. [35] as,

Primal Problem

$$\begin{aligned} \text{maximize : } & \frac{1}{2} \text{Tr } \hat{X} + \frac{1}{2} \text{Tr } \hat{X}^\dagger \\ \text{subject to : } & \begin{pmatrix} \mathcal{E}(\hat{\rho}) & \hat{X} \\ \hat{X}^\dagger & \Phi(\hat{\sigma}) \end{pmatrix} \geq 0 \\ & \hat{X}, \hat{\sigma}, \hat{\rho} \in \mathcal{P} \end{aligned}$$

Dual Problem

$$\begin{aligned} \text{minimize : } & \max_{\vec{u} \in \mathbb{C}^n, |\vec{u}| \leq 1} \sqrt{\text{Tr} [\vec{u}^\dagger \hat{Y}^\dagger \hat{Y} \vec{u}]} \\ \text{subject to : } & \hat{Y} \text{ must be positive definite} \end{aligned}$$

The primal problem is easier to solve. It can be converted into a linear program which is a convex optimization problem [27] with scalar constraints instead of the matrix above that has guaranteed global minima i.e. any local optimal solution is a global optimal solution. These can be found using the method of interior points which converts the problem into finding iteratable feasible solutions that lives in a geometric region defined by hyperplane intersections of constraints provided above.

2.7 Scaling laws for minimax-estimator risk with ensemble size

The Dnorm error is an example of a worst-case measure of error or risk. The QST method used in SQPT to reconstruct $\hat{\rho}'$ is done by taking the mean of the various samples drawn from measurement distributions after applying the channel on input state $\{\hat{\rho}\}$. These are independent random variables⁸ and tend to a Gaussian in the large measurements-size limit following the central limit theorem. The mean of Gaussians is a *minimax*, the best worst-case, error estimator [37]. Then, building on decision theory framework, it has been shown that measures of worst-case error or risk scale as $O(1/\sqrt{N})$ for N number of measurements for minimax estimators in the QST method that are fixed during execution. This is a quantum-specific result as classical probability estimation scales as $O(1/N)!$ Methods that are online have their risks scaled as $O(1/\exp\{N\})$. The exact argument construction in a Bayesian-Likelihood information-theoretic framework can be read in the PhD thesis [1].

⁸multinomial, see [Section 3.2](#)

Chapter 3

Computational standard QPT model

In this chapter, two computational methods of doing QPT simulations will be presented. The first, via frequentist maximum likelihood, in [Section 3.1](#) is unstable and its pitfalls will be analysed to propose another simpler method in [Section 3.2](#). This one uses a linear inversion map applied on simulated experimental outcomes, is regularized, and converges. It also allows the input probe states to be parametrized. Classical noise models to simulate imperfections are discussed in [Section 3.3](#). A performance scheme is constructed for these parametrizations in [Section 3.5](#).

3.1 Maximum likelihood method: initial approach and pitfalls

Following Ziman et. al. [38, 39] and assuming the general form of a CPTP channel as that performing an affine transformation on $\hat{\rho}(\vec{r})$ ([2.22](#)),

$$\mathcal{E}(\hat{\rho}(\vec{r})) = \begin{pmatrix} 1 & \vec{0} \\ \vec{c} & \hat{M} \end{pmatrix} \begin{pmatrix} 1 \\ \vec{r} \end{pmatrix} = \begin{pmatrix} 1 \\ \hat{M}\vec{r} + \vec{c} \end{pmatrix} \quad (3.1)$$

where the vectors are 3 dimensional e.g. $\vec{c} = (c_x, c_y, c_z)$. Spectrally decomposing or diagonalizing $\hat{M} \rightarrow \hat{U}\hat{M}\hat{U}^\dagger$ and ignoring the unitaries \hat{U} that parametrize the orientation of each axis that M transforms, a single qubit channel map \mathcal{E} has 6 degrees of freedom without loss of much physical generality (if the map is an ellipsoid in Bloch space, then we just lose relative orientation). \mathcal{E} can be rewritten in superoperator form,

$$\bar{\mathcal{E}}(\hat{M}, \vec{c}) = \begin{pmatrix} 1 + M_z + c_z & 0 & 0 & 1 - M_z + c_z \\ c_x + ic_y & M_x + M_y & M_x - M_y & c_x + ic_y \\ c_x - ic_y & M_x - M_y & M_x + M_y & c_x - ic_y \\ 1 - M_z + c_z & 0 & 0 & 1 + M_z + c_z \end{pmatrix}$$

Converting to its choi state by partial transposition, below 2 inequalities must be obeyed for it to be a) positive definite and b) trace non-increasing [[40](#)],

$$(1 \pm M_x)^2 \geq (M_y \pm M_z)^2$$

$$|c_i| + |M_i| \leq 1 \quad \forall i \quad (3.2)$$

Now channel elements can be directly related with transformed state elements $\vec{r} = \hat{M}\vec{r} + \vec{c}$. Then consider the 6 eigenstates $\{| \pm \rangle_x, | \pm \rangle_y, |0/1\rangle\}$ of the Pauli matrices as input probes. And their corresponding pure states of the form $\{\hat{\rho}_{\pm}^x, \hat{\rho}_{\pm}^y, \hat{\rho}_{\pm}^z\} = \{| \pm \rangle \langle \pm |_x, | \pm \rangle \langle \pm |_y, |0/1\rangle \langle 0/1|_z\}$ as measurement operators. One can decompose a Pauli-operator measurement (4.1),

$$\text{Tr } \hat{\sigma}_i \hat{\rho}_j = \text{Tr } [\hat{\rho}_+^x \hat{\rho}_j - \hat{\rho}_-^x \hat{\rho}_j] = \text{Tr } \hat{\rho}_+^x \hat{\rho}_j - \text{Tr } \hat{\rho}_-^x \hat{\rho}_j = +1p(+^x|\rho_j) + -1p(-^x|\rho_j) \quad (3.3)$$

using its rank 1 decomposition, trace linearity and lastly postulate 3. Note that $p(+^x|\rho_j) = 1 - p(-^x|\rho_j)$ trivially due to binary outcomes. Using (3.3), we get a set of true binomial probability vectors $\{\vec{p}^x\}$ e.g. $\vec{p}^x = [p(+^x), p(-^x)]$ that decompose the transformed state into a set of measurement outcomes from applying Pauli operators to simulate experimental settings (e.g. like photon polarization slit experiments [26, 41]). Then, random sampling algorithms like `numpy.binomial` can simulate outcomes \vec{P}^x of M measurements by sampling $\vec{p}^x M$ times and renormalizing to get $\vec{p}_{\text{est}}^x = \vec{P}^x/M$. Then, the likelihood function for the M measurements experiment is given by,

$$L(\mathcal{E}) = -\log \prod_{k=1}^M p(k|k) = -\sum_{k=1}^M \log \text{Tr } \mathcal{E}(\hat{\rho}_k) \hat{F}_k = -\sum_{p=1}^6 a_p \log \text{Tr } \mathcal{E}(\hat{\rho}_p) \hat{F}_p \quad (3.4)$$

where $\hat{F}_k \equiv \hat{\rho}_{\pm}^i$ stores the outcome k as a measurement operator. We can also collect all the outcomes for each eigenstate and just add a multiplying constant $a_p = P_{\pm}^i/M$ to get a computationally friendly (3.4). Regularizing the estimated channel as,

$$\mathcal{E}'_{\text{est}} = (1-k)\mathcal{E}_{\text{est}} + k\frac{1}{2}$$

so that the estimator remains CP, we can maximize (3.4) to obtain the set of 7 parameters $\vec{O} = [\vec{M}, \vec{c}, k]$ subject to the constraints (3.2). I used a standard constrained optimization by linear approximation (COBLYA) in `scipy.optimize.minimize` that iterates greedily through local solutions from approximating the constrained problem as a linear program (Section 2.6). After this, I just evaluated the l1-norm $\sum_i |O^i - O_{\text{est}}^i|$ to see how well the scheme did as this error upper bounds all other channel errors.

This was my first approach to simulating a QPT protocol. There were a few issues with this:

- Random sampling aside, there was an unacceptable amount of variation in the $\mathcal{E}'_{\text{est}}$ parameters due to local optima not being the same as global optima in (3.4). This was an endemic problem and could not be ruled out even with more customized QST maximization schemes like diluted iterations, non-convex methods [42–44] that guarantee minimization at every step. I would have to repeat the procedure under identical conditions for many M -ensemble sizes and the problem of many local minimas due to unitary freedom of channels would not fix the end point of these improved methods.
- This method was not general and ruled out the most basic channel: the unitary rotation, which would prove very useful in debugging and understanding future schemes.

- Despite regularization, the estimator is biased and always underestimates [45].

3.2 Naturally regularized SQPT model with fixed measurement design

Learning from past failures, I decided to formulate a *general* QPT scheme. This would rely on estimating the *generalized* Krauss operators and a channel of the form,

$$\mathcal{E}(\cdot) = \sum_i \sum_j \chi_{ij} \hat{E}_i(\cdot) \hat{E}_j^\dagger \quad (3.5)$$

that essentially is a computational version of SQPT mentioned in [Section 2.4](#) with the familiar X matrix. At its heart, lies the notion (inspired by postulate 4) that an n -qubit basis can be written as

$$\begin{aligned} \{\mathbb{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}^{\otimes n} &= \{\mathbb{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\} \otimes \{\mathbb{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\} \otimes \{\mathbb{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}^{\otimes(n-2)} \\ &= \{\mathbb{1} \otimes \mathbb{1}, \mathbb{1} \otimes \hat{\sigma}_x, \mathbb{1} \otimes \hat{\sigma}_y, \mathbb{1} \otimes \hat{\sigma}_z, \otimes \cdots \otimes \hat{\sigma}_z \otimes \hat{\sigma}_z\} \otimes \{\mathbb{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}^{\otimes(n-2)} \end{aligned}$$

which can be defined recursively on a computer and are orthogonal under the Hilbert-Schmidt inner product ([2.37](#)) and \hat{E}_i, \hat{E}_j belong to this basis. I avoid the final diagonalization step to bring \mathcal{E} into operator-sum form, just to avoid an extra source of variation. Thus, the QPT problem has been translated into simply estimating the coefficients χ_{ij} that go with these basis elements.

Now we can decompose each $\hat{\sigma}_i$ into its rank 1 eigenstates following ([3.3](#)). Expanding this out for each qubit after taking the trace and applying the tensored measurement operators gives us *multinomial* probability vectors. This basically means that instead of 2 there are now multiple outcomes that come with each measurement operator application. For example, take $\hat{\rho} = \hat{\rho}_A \otimes \hat{\rho}_B$ which has 16 orthogonal QST basis decompositions ([2.28](#)). One of these is the basis $\hat{\sigma}_z \otimes \hat{\sigma}_z$. We can repeat the trick in ([3.3](#)),

$$\begin{aligned} \text{Tr } \hat{\sigma}_z \otimes \hat{\sigma}_z \hat{\rho} &= \text{Tr } [(\hat{\rho}_+^z - \hat{\rho}_-^z \otimes \hat{\rho}_+^z - \hat{\rho}_-^z) \hat{\rho}] \\ &= \text{Tr } [(\hat{\rho}_+^z \otimes \hat{\rho}_+^z - \hat{\rho}_+^z \otimes \hat{\rho}_-^z - \hat{\rho}_-^z \otimes \hat{\rho}_+^z + \hat{\rho}_-^z \otimes \hat{\rho}_-^z) \hat{\rho}] \\ &= \text{Tr } [\hat{\rho}_+^z \otimes \hat{\rho}_+^z \hat{\rho}] - \text{Tr } [\hat{\rho}_+^z \otimes \hat{\rho}_-^z \hat{\rho}] - \text{Tr } [\hat{\rho}_-^z \otimes \hat{\rho}_+^z \hat{\rho}] + \text{Tr } [\hat{\rho}_-^z \otimes \hat{\rho}_-^z \hat{\rho}] \\ &= p(+^z +^z | \hat{\rho}) - p(+^z -^z | \hat{\rho}) - p(-^z +^z | \hat{\rho}) + p(-^z -^z | \hat{\rho}) \end{aligned}$$

There are now 4 outcomes. After, ignoring and storing the algebra of their eigenvalues separately, one arrives at a multi-outcome probability vector $\vec{p}^{z_1 z_2}$! The whole set $\{\vec{p}\}$ can be randomly sampled using `numpy.multinomial` to perform M -ensemble QST on the transformed $\hat{\rho}' = \mathcal{E}(\hat{\rho})$ using the same procedure as before. Note that the rank 1 projectors of $\hat{\mathbb{1}}$ have degenerate eigenvalues of +1 and reveals no information about the state. As a potential measurement operator, it can thus be ignored in whatever measurement set I will henceforth discuss. Also from now on, I will fix the Pauli basis for measurements and call it a *fixed measurement design*. It is naturally regularized: as long we ensure reconstructed $\hat{\rho}_{\text{est}}$ is physical, the channel is automatically CPTP. Coming to the QPT step, to be able to estimate every single

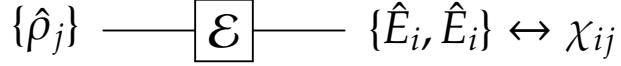


Figure 3.1: SQPT circuit for input state set $\{\hat{\rho}_j\}$ and reconstructed fixed Krauss operators from the χ matrix.

element of an arbitrary $\hat{\rho}$, we have to choose a set of input states or *preparation design*, apply \mathcal{E} and reconstructing X by stacking the vectorized outcomes $[\vec{\rho}']$ as $|\lambda\rangle$ and applying the inverse encoding matrix $\bar{\mathbf{B}}^{-1}$ to get $X = \bar{\mathbf{B}}^{-1} |\lambda\rangle$. The most general form of arbitrary elements of $|iXj|$ of $\hat{\rho}$ can be written as,

$$\hat{\rho}_{ij} = |i\rangle\langle j| = \alpha |a\rangle\langle a| + \beta |b\rangle\langle b| + \gamma |c\rangle\langle c| + \delta |d\rangle\langle d| \quad (3.6)$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{C}$ are arbitrary coefficients that can be fixed by choosing the pure physical states a, b, c, d e.g. using polar qubits (2.4) and absorbing global phase factors into the coefficients. Generally, we can choose as many physical states as we like but the minimum is 4 due to the span of the Pauli operator space as seen from the qubit parametrization equation. We solve (3.6) through an unconstrained optimization via sequential linear programming (SLQP) by minimizing the sum of complex and real component-wise differences between matrix elements in the LHS and RHS. This is just iterations over local solutions parameter space until an acceptable accuracy is obtained. A tolerance of 10^{-8} was used as a benchmark in this process as on average the orthogonal parametrization (2.31), which has an analytical form, averaged no better and was used as a fiducial marker with which all other parametrizations were compared. Floating-point errors are an artefact of the reals being modelled by a finite set of points such that in a computer there is a one-one map: $[0, \infty) \rightarrow [0, 1]$. The smallest number, the mantissa, for my machine was 10^{-16} was much smaller than the tolerance that yielded physics of appropriate resolution. So, without divulging special treatment to floating point arithmetic, simply keeping the order of magnitudes of numerics far above it was enough. Finally, $D_{\text{norm}} \|\mathcal{E}_{\text{est}} - \mathcal{E}\|_\diamond$ was computed by solving the primal program using the cvxpy library stated in Section 2.6).

3.3 Noise model

The measurement operators $\hat{\rho}^i$ can be modified to incorporate imperfections that are common in realistic situations. Realistically, this noise is sourced from a large reservoir of source variables and a common practice is to use white or Gaussian noise. However, to keep things simple and avoid complicated measurement statistics, I decided to use a constant, linear noise model \vec{e} ,

$$\hat{\rho}_{\pm}^i = (1 - \epsilon^i) \hat{\rho}_{\pm}^i + \epsilon^i \frac{1}{2} \hat{\mathbb{I}} \quad (3.7)$$

for some axis $i = x, y, z$ or all. Here, $\epsilon^i \in [0, 1] \in \mathbb{R}$ and is not a random variable. This is the same as the depolarizing channel (2.25) and the effect is simply to decohere the measurement outcomes. Hence, this is decohering noise. But we would like the noise to converge in the large measurements limit. By taking the trace on both sides

3 Computational standard QPT model

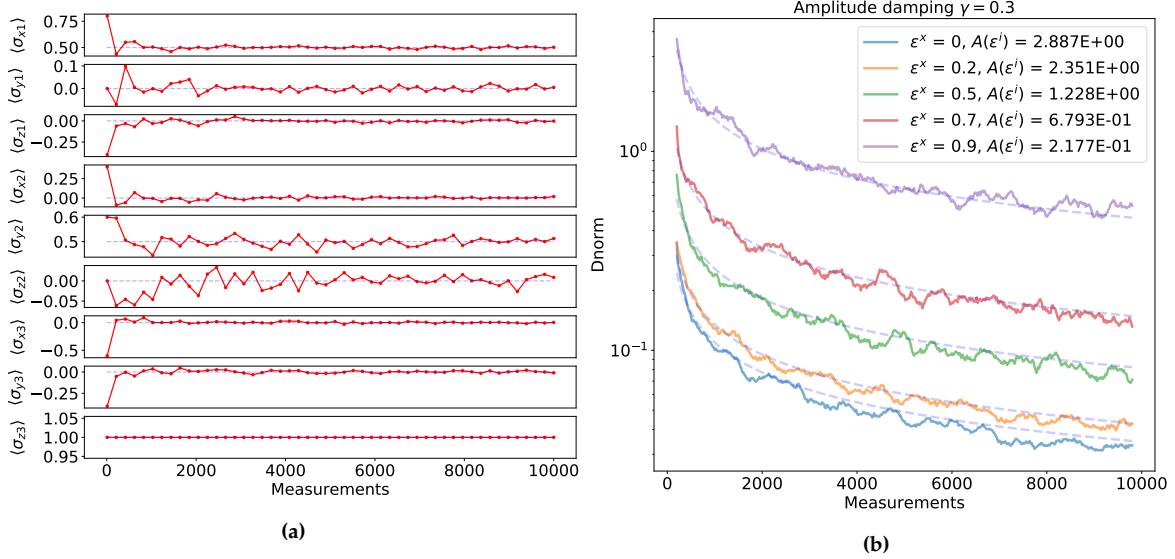


Figure 3.2: In a) we have QST for 3-axis Pauli measurements for the Pauli eigenstates $0.5|+\rangle\langle+$, $0.5|+\rangle\langle+$, $|0\rangle\langle 0|_z$ converging to their respective eigenvalues $+0.5$, $+0.5$, $+1$ when the measurements become large. In b) The D_{norm} vs 3-axis measurements decay for various x axis noise levels ϵ^x is shown. The tomographed channel is $\mathcal{D}(\gamma = 0.4)$. The Power law ansatz is fit and the performance factors $A(\epsilon^i)$ are computed which decrease with noise. Here all axes get the same number of measurements in steps of 10 from 10 to 10000. A moving average has been applied to make the trend more visible.

of (3.7) and rearranging, we can apply a correction,

$$\text{Tr } \hat{\rho}_{\pm}^i \hat{\rho} = \frac{\text{Tr } \hat{\rho}_{\pm}^i \hat{\rho} - \text{Tr } \epsilon_T^i \hat{\rho}}{1 - \epsilon_T^i} = \frac{\text{Tr } \hat{\rho}_{\pm}^i \hat{\rho} - \epsilon_T^i}{1 - \epsilon_T^i} \quad (3.8)$$

This would be done *after* sampling from the noisy multinomial probability vector to get closer to the “true” probability in QST. Technically, the proper thing to do would be to parametrize the noise $\vec{\epsilon}$ as a function of measurement and construct a probability distribution that converges to a constant value ϵ_T^i . However, that would be another project and in the interest of saving time, I just set $\epsilon_T^i = \epsilon^i$ as my objective was to add variation in small M -ensembles.

In retrospect, this model was qualitative at best and did not have a high enough resolution or power to exceed quantum fluctuations sourced from multinomial sampling in the QST step (see Section 4.6). I tried modelling $\vec{\epsilon}$ as Binomial, Gaussian and Poisson random variables as well but the effects on CQPT were harder to interpret and were still weak compared with quantum fluctuations.

3.4 Tests

I will recall and collate a few crucial unit tests that were mentioned in passing until now to ensure that the computational SQPT protocol was working as expected:

- That the reconstructed states $\hat{\rho}'$ were physical, noisy or not, was ensured by computing explicitly and confirming $\text{Tr } \hat{\rho}' = 1$ and states were renormalized, if necessary.

- Likewise the choi state $\hat{\rho}_{\mathcal{E}}$ of the reconstructed channel $\{\hat{E}_i, \hat{E}_j\}$ was checked for a positive eigenvalue spectrum that is necessary for a CP map.
- A number of example states $\hat{\rho}$ and channels \mathcal{E} served as diagnostic checks and were reproduced by the QST and QPT protocols.
- The convergence of probabilistic fluctuations from random sampling and subsequently in the Dnorm is shown in figure 3.2 for various noise levels ϵ^i .

3.5 Power law fit for performance evaluation

Following the argument in [Section 2.7](#), We can construct a power law fit for the Dnorm $\|\mathcal{E}_{\text{est}} - \mathcal{E}\|_{\diamond}$ as a function of M , the measurement ensemble size and noise ϵ^i

$$\text{Dnorm}(M, \epsilon^i) = \frac{1000}{A(\epsilon^i)} M^{\frac{1}{2}} \quad (3.9)$$

where $A(\epsilon^i)$ is a coefficient that signifies the rate of decay of the Dnorm. Greater $A(\epsilon^i)$ correspond to faster decays and hence fewer need to make measurements to reduce errors below a certain level. Note that due asymptotic equivalence of noisy matrix completion and quantum state tomography (QST) [13], in the limit of large measurement size, the random sampling from different multinomial probability vectors tends to become Gaussian and the standard deviation argument for the variation in the $\hat{\rho}_{ij}$ matrix elements as the source of error is also sensible. By repeating in steps of 10 for measurement sizes M till 10,000, a log-log linear regression fit shows strong agreement with both arguments for the Dnorm decay with M . Adding a constant noise term simply translates the curve upwards and decreases $A(\epsilon^i)$. This is shown in figure 3.2. A summary of the computational SQPT algorithm can be found in [algorithm 1](#).

3 Computational standard QPT model

Algorithm 1: Noisy computational SQPT performance model

```

1 function: noisy QST ( $n, \hat{\rho}, \vec{\epsilon}, M$ );
  Input :  $n$ -qubit state  $\hat{\rho}$ , 3-axis noise vector  $\vec{\epsilon}$ , number of measurements  $M$ 
  Output: noisy probability vector set  $\{\vec{p}\}$  of measurement operators  $\hat{\rho}_{\pm}^{i \otimes n}$ 
2 Decompose Pauli operators  $\hat{\sigma}_i^{\otimes n}$  into rank 1 eigenstates  $\hat{\rho}_{\pm}^{i \otimes n}$  (3.3)
3 Make the modification  $\hat{\rho}_{\pm}^{i \otimes n} \xrightarrow{\epsilon^i} \hat{\rho}_{\pm}^{i \otimes n}$  to add constant noise
4 for  $i$  in  $\{\hat{\rho}_{\pm}^{i \otimes n}\}$  do
  5 Construct multinomial probability vector  $\vec{p}^i$  of outcomes
  6 Random sample  $M$  times from  $\vec{p}^i$  to get vector of outcomes  $\vec{P}^i$ 
  7 Normalize to get  $\vec{p}_{\text{est}}^i \leftarrow \vec{P}^i/M$ 
  8 Apply noise correction  $\vec{p}'_{\text{est}} \xleftarrow{(3.8)} \vec{p}_{\text{est}}^i$ 
9 end
10 function: QPT ( $n, \{\hat{\rho}_j\}, \vec{\epsilon}, M, \mathcal{E}, t$ );
  Input : new are  $n$ -qubit preparation design  $\{\hat{\rho}_j\}$ , CPTP channel  $\mathcal{E}$ , tolerance
     $t = 10^{-8}$ 
  Output: Dnorm  $\|\mathcal{E}_{\text{est}} - \mathcal{E}\|_{\diamond}$ , estimated channel  $\mathcal{E}_{\text{est}}$  in form  $\{\hat{E}_i, \hat{E}_j\}$ 
11 First fix coefficients of  $\{\hat{\rho}_j\}$  needed for estimating  $\hat{\rho}'_{ij} \xleftarrow{\mathcal{E}} \hat{\rho}_{ij} = |i\rangle\langle j|$ 
12 for  $i, j$  in  $|i\rangle\langle j|$  do
  13 Use preparation design  $\{\hat{\rho}_j\}$  to solve the general  $\hat{\rho}_{ij}$  parametrization (3.6)
    by the SLQP method for fixed upper limit of iterations and output  $t$ 
  14 if  $t < 10^{-8}$  then exit protocol and raise flag
15 end
16 Apply channel on preparation design  $\{\hat{\rho}_j\} \xrightarrow{\mathcal{E}} \{\hat{\rho}'_j\}$ 
17 for  $j$  in  $\{\hat{\rho}'_j\}$  do
  18 | noisy QST to produce noisy transformed states  $\hat{\rho}'_{j,\text{est}}$  for  $M, \vec{\epsilon}$ 
19 end
20 Vectorize  $\{\hat{\rho}'_j\}_{\text{est}} \rightarrow |\lambda\rangle$ 
21 Construct  $\bar{\mathbf{B}}$  using generalized basis (3.5) with elements  $B_{lm,ij} = \hat{E}_l |i\rangle\langle j| \hat{E}_m^\dagger$ 
22 Get  $|\lambda\rangle = \bar{\mathbf{B}}^{-1} |\lambda\rangle$  through matrix inversion
23 Get  $\mathcal{E}_{\text{est}}$  using  $\lambda$ 
24 Compute Dnorm  $\|\mathcal{E}_{\text{est}} - \mathcal{E}\|_{\diamond}$  by defining the triple and solving the primal
    problem using linear programming approximations to the semi-definite
    program (sec. 2.6)
25 function: performance ( $\{M\}, \|\mathcal{E}_{\text{est}} - \mathcal{E}\|_{\diamond}^{(M)}$ );
  Input : new are  $M$ -ensemble array  $\{M\}_{s=10}^{10,000} = \{10, 20, \dots, 10000\}$ , with
    associated Dnorm  $\|\mathcal{E}_{\text{est}} - \mathcal{E}\|_{\diamond}^{(M)}$ )
  Output:  $A(\epsilon^i)$ , confidence intervals
26 Convert to log space for  $\|\mathcal{E}_{\text{est}} - \mathcal{E}\|_{\diamond}^{(M)}$  values and use least squares
    minimization with the Power law ansatz (3.9) with discrete points  $\{M\}_{s=10}^{10,000}$ 
27 Compute  $1000/A(\epsilon^i)$  as the y-intercept

```

Chapter 4

Results: varying preparation designs

In this chapter, using formalism developed in [Chapter 2](#) and the computational SQPT model from [Chapter 3](#), varying preparation designs is studied, explicitly, for the single qubit case. Variation is treated as an optimization problem. There are essentially 4 degrees of freedom that can be varied for the 1-qubit preparation design implying that one must prepare at least 4 different initial state ensembles to perform a *complete* QPT. This is because the linearly-independent¹ vector set that specifies a 1-qubit has 4 elements in it,

$$\text{span}(\hat{\sigma}_\alpha) = \{\hat{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\} \quad (4.1)$$

Here, I present my results in order that I obtained them. My approach has mainly been:

- Unfix 1 d.o.f. at a time.
- Exploit some underlying symmetry in the parametrization of the d.o.f. and do a discretized grid search of the parameter space.
- Analyze the effects on QPT performance factors $A(\epsilon^i)$ of this variation under various noise levels ϵ^i .

This culminates into discussions in [Section 4.2](#) and [Section 4.3](#). However, this approach was not feasible for more than 2 d.o.fs for two reasons. Not only was it not computationally feasible to explore an exponentially growing parameter space, it was also difficult to manually analyze the nuanced higher dimensional landscape of performance factors due to inherent quantum and decohering noise model fluctuations. It turned out that this was not a big issue as by [Section 4.3](#), it was becoming clearer that the optimal preparation designs would be symmetric. So, I decided to directly investigate symmetric configurations in [Section 4.5](#) and found the tetrahedral preparation design to be optimal under the performance framework for all noise levels. Distributing measurements in [Section 4.6](#) under noise-asymmetry had limited to no success for unitary and non-unitary channels.

¹SU(2) operator

4.1 Noisy channels

Let's begin by considering actions of the 1-qubit channels I have mentioned in [Chapter 2](#) with an addition of decohering noise. Working in the Bloch vector representation of the 1-qubit density matrix $\hat{\rho}$ in [\(2.8\)](#),

$$\hat{\rho}(\vec{r}) = \frac{1}{2} (\hat{\mathbb{1}} + \vec{r} \cdot \vec{\sigma}) = \frac{1}{2} (\hat{\mathbb{1}} + r_x \hat{\sigma}_x + r_y \hat{\sigma}_y + r_z \hat{\sigma}_z)$$

In case of unitary transformations, which we saw were equivalent to rotations of \vec{r} , under decohering noise, the transformation can become asymmetric,

$$\vec{r} \xrightarrow{\hat{R}} \vec{r}' \xrightarrow{+\vec{\epsilon}} \vec{r}'(1 - \vec{\epsilon})$$

if the axis noise vector $\vec{\epsilon}$ has differing values for at least 2 different axes. By virtue of QST [\(2.28\)](#), the asymptotic Pauli measurements in the large ensemble-size limit are just reconstructions of the transformed Bloch vector components

$$\langle \hat{\sigma}_j \rangle \equiv \text{Tr} (\hat{\sigma}_j \hat{\rho}') = r'_j (1 - \epsilon_j) \quad (4.2)$$

under rotations. Hence, the noise model can introduce asymmetries to otherwise symmetric channel operations.

Consider now the form of the reconstructed components $\langle \hat{\sigma}_j \rangle$ under a symmetric non-unitary channel operation like the depolarizing channel $\mathcal{P}(p)$ [\(2.25\)](#),

$$\langle \hat{\sigma}_j \rangle = r_j \left(1 - \frac{2}{3}p\right) (1 - \epsilon_j) = r'_j (1 - \epsilon_j) \quad (4.3)$$

And for the asymmetric amplitude damping channel $\mathcal{D}(\gamma)$

$$\langle \hat{\sigma}_j \rangle = \begin{cases} (1 - \epsilon_j)r_j \sqrt{(1 - \gamma)} & \text{if } j = x \text{ or } y \\ (1 - \epsilon_j)r_j(1 - \gamma) & \text{if } j = z \end{cases} \quad (4.4)$$

And for the asymmetric dephasing channel $\mathcal{H}(p)$ [\(2.26\)](#),

$$\langle \hat{\sigma}_j \rangle = \begin{cases} (1 - \epsilon_j)r_j & \text{if axis is invariant} \\ (1 - \epsilon_j)r_j p & \text{if axis is not invariant and will be dephased} \end{cases} \quad (4.5)$$

Thus, the channels are transformed non trivially under axis noise. Let us again consider the action of these *noisy* channels on the whole Bloch sphere. For example, the depolarizing channel which was initially symmetric in its action on all 3 axes loses the x-axis symmetry when $\vec{\epsilon} = (\epsilon^x, 0, 0)$ and becomes similar to the dephasing channel. Similarly, both the dephasing and amplitude damping channels are symmetric in two axes and introduction of noise in any breaks this symmetry. This will be important when considering optimal initial states for QPT reconstruction as we need to find a way to *symmetrize* the distribution of noise including quantum fluctuations due to finite-size ensembles. What does this mean in terms of varying preparation design?

From, [\(2.31\)](#), we know that the SQPT configuration is inherently asymmetric

in terms of reconstruction where the diagonal components ($i = j$) of $\hat{\rho}$ are measured in the Z-eigenbasis and the off diagonals ($i \neq j$) are computed following the prescription,

$$|i\rangle\langle j| = |+\rangle\langle +|_x + i|-\rangle\langle -|_y - \frac{1+i}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \quad (4.6)$$

. This, as it will turn out, is not optimal in terms of the heuristic noise distribution argument made above. To computationally motivate this intuition and attempt to answer objective 1 of this thesis (Section 1.2), I will consider varying single and double d.o.fs in (4.6) while fixing the computational basis in the proceeding two sections.

4.2 Single axis rotations

I will start by considering (4.6) following [6].

Rotation operators of the form, $\hat{R}(\theta, \vec{n}) = e^{-i\frac{\theta}{2}\vec{n}\cdot\vec{\sigma}}$, can be applied on the first term as $\hat{R}|+\rangle\langle +|_x\hat{R}^\dagger$ while fixing the remaining coefficients through the convex optimization routine (1). Then, the d.o.f. can be varied by generating a grid of rotation operators. However, this procedure can be simplified by noting that it is equivalent to simply discretizing the Bloch sphere and picking values of the first term that correspond to the lattice points on the sphere i.e.

$$\hat{R}(\theta, \vec{n})|+\rangle\langle +|_x\hat{R}^\dagger(\theta, \vec{n}) \leftrightarrow |\psi(\theta_1, \phi_1)\rangle\langle\psi(\theta_1, \phi_1)|$$

where (θ_1, ϕ_1) are polar angles that are defined using the convention shown in figure 4.1a. This is completely equivalent to the Bloch vector \vec{r} as this is normalized and has effectively only 2 d.o.fs for pure states. By noting that \vec{r} for the four terms in (4.6) in that order are,

$$\vec{r}_x = (1, 0, 0) \quad \vec{r}_y = (0, 1, 0) \quad \vec{r}_{z_1} = (0, 0, 1) \quad \vec{r}_{z_2} = (0, 0, -1)$$

where I index the vectors by the eigenstates that they represent. It should be apparent that the states represent 3 orthogonal axes that parametrize the preparation design. Realizing that only the relative angle between each axis is important will allow a reduction in the size of parameter space, working in polar co-ordinates (see figure 4.1a),

$$[0, \pi)_{\theta_x}, [0, 2\pi)_{\phi_x} \rightarrow [0, \pi/2)_{\theta_x}, [0, \pi)_{\phi_x} \quad (4.7)$$

I deliberately conducted the grid-sweep on the bigger interval first before reducing the parameter space to confirm that this was the case for the first run of data collection. The parameter space was discretized into 20×20 steps per angle. Figure 4.1b displays the results of varying \vec{r}_x for SQPT of the amplitude damping channel $\mathcal{D}(\gamma = 0.2)$ where the other 3 vectors were fixed with noise $\vec{\epsilon} = (0.5, 0, 0)$ at the measurement stage. The contours are smoothly increasing as the angular sweep approaches relative orthogonality with the other 3 initial states. The peaks are located at $(\theta_1, \phi_1) = (\frac{\pi}{2}, 0), (\frac{\pi}{2}, \pi), (\frac{\pi}{2}, 2\pi) \leftrightarrow |+\rangle_x, |-\rangle_x, |+\rangle_x$ despite the fact that the x-axis is noisy.

Repeating the variation procedure for lower noise levels replicates the same results but with higher $A(\epsilon^i)$ values which I was pleased to confirm, as it was an

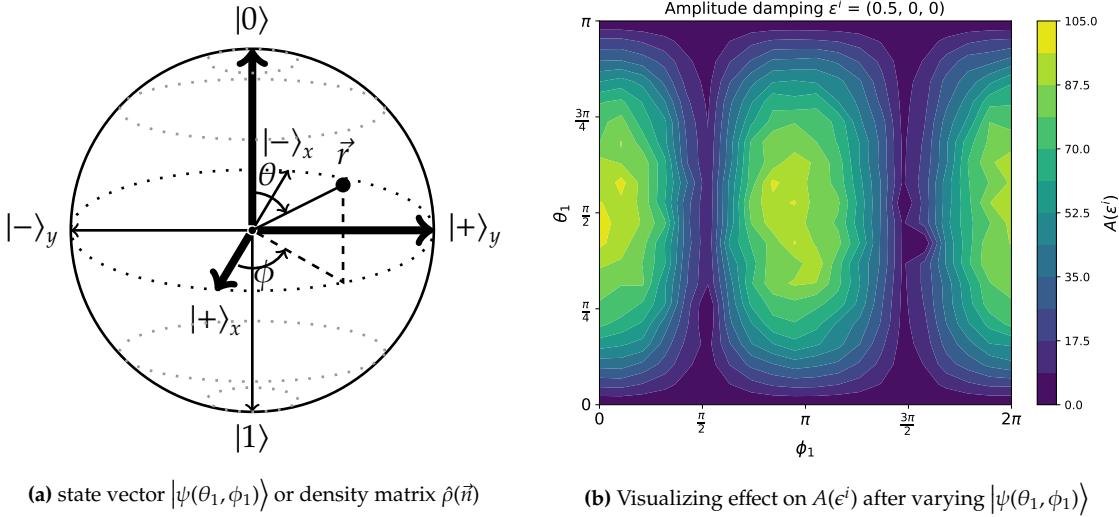


Figure 4.1: In a) is a geometric representation of the Bloch vector axes with corresponding eigenstates. An arbitrary pure 1-qubit represented by polar angles (θ_1, ϕ_1) is shown (equivalently \vec{r}). Note the great circles parallel to and on the equatorial plane which proxy for my discretization method for points on the Bloch sphere. The reduced parameter space (4.7) is highlighted by the quadrant enclosed by thick lines. In b), a 2-D contour plot can be seen of the performance factors $A(\epsilon^i)$ vs (θ_1, ϕ_1) variation of the x-eigenstate in (4.6). The shape is parabolic and peaks at $(\phi_1 = 0, \pi, 2\pi, \theta_1 = \frac{\pi}{2})$ coinciding with $|+\rangle_x, |-\rangle_x, |+\rangle_y, |+\rangle_x$, all orthogonally lying on the axes in a).

expected consequence of the constant noise model chosen in [Chapter 3](#). The same was also true for other symmetric channels like unitary rotations and depolarization (see figure 4.2).

4.3 Dual axis rotations

Repeating the previous prescription with an additional d.o.f, the initial state parametrization is now,

$$\hat{\rho}_{ij} = |i\rangle\langle j| = \alpha |\psi(\theta_1, \phi_1)\rangle\langle\psi(\theta_1, \phi_1)| + \beta |\psi(\theta_2, \phi_2)\rangle\langle\psi(\theta_2, \phi_2)| + \gamma |0\rangle\langle 0| + \delta |1\rangle\langle 1| \quad (4.8)$$

This effectively doubles the size of the parameter space that needs to be swept and some computational optimizations are in order. Firstly, working with the reduced parameter space (4.7) for grid sweeping is valid as evidenced by previous computational arguments. The number of configurations to be considered is still $10 \times 10 \times 10 \times 10$. However, realizing that interchanging $\theta_1, \phi_1 \leftrightarrow \theta_2, \phi_2$ is a symmetry reduces this by half. Also, there are certain configurations that don't mathematically add up to the specific $\hat{\rho}_{ij}$ element. This can be confirmed by performing the convex procedure for computing coefficients for that specific design (4.8) before the QPT step.

This significantly reduces the number of configurations that need to be considered to be strictly less than 5000. However, I had to consider computing $A(\epsilon^i)$ vs (θ, ϕ) landscapes for 6 channels (x, y, z rotations and non-unitary $\mathcal{P}, \mathcal{D}, \mathcal{H}$) at 3 noise levels ($\epsilon^x = (0, 0.3, 0.5)$). Only symmetry-breaking noise for these seemed useful as the rest of the noisy channels just morphed into these channels. But there was still a bigger issue to consider. One plot in figure 4.2 was estimated to take about

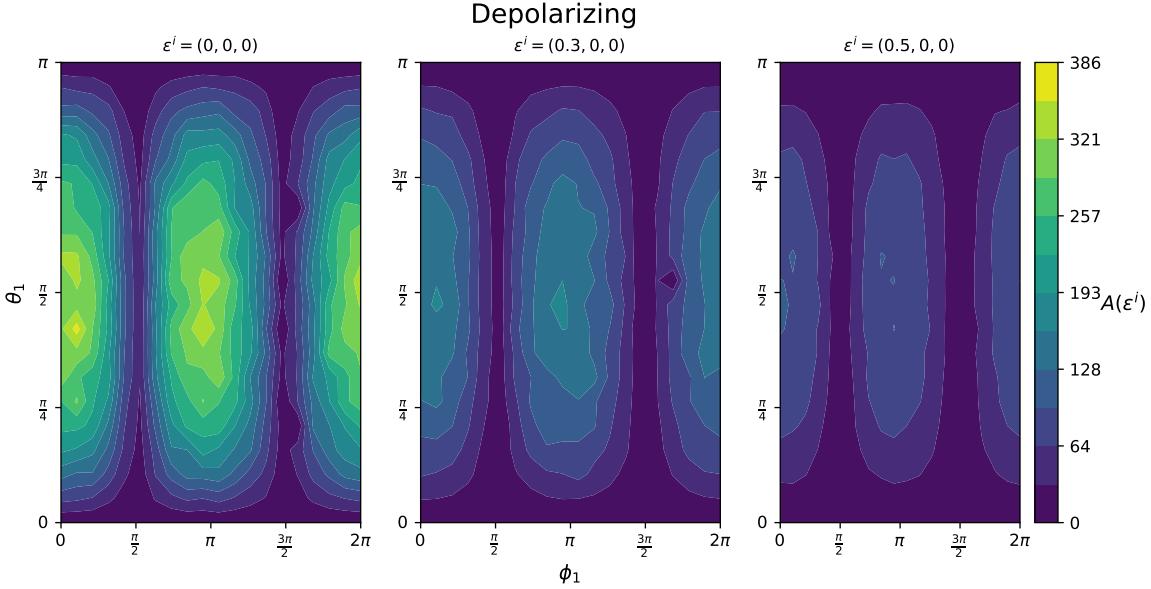


Figure 4.2: Variation of QPT performance $A(\epsilon^i)$ due to rotations of the x-eigenstate (\vec{r}_x) in (4.6)) for the symmetric depolarizing channel $P(\alpha = 0.6)$. As the noise ϵ^x is just constant, the effect is a contracting/scaling of $A(\epsilon^i)$. There are 13 levels in the contour plot above which are defined by the leftmost non-noisy case and the landscape shape is still the same for the noisy cases $\epsilon^x = 0.3, 0.5$, but with masked resolution of lower levels. The peaks can again be seen at $(\phi_1 = 0, \pi, 2\pi, \theta_1 = \frac{\pi}{2})$. The redundancy in results outside the reduced parameter space (4.7) should be noted.

10 minutes (reduced space): hence, the total compute time for the entire variation procedure was upper bounded by $12 \times \frac{1}{6 \times 24} \times 500 \approx 42$ days.

This prompted me to consider parallelizing computation and batch distributing it on the College's computing cluster [46]. A straightforward solution as the complexity arose from 2 nested loops over the parameter space: one with respect to θ_1, ϕ_1 and the other due to θ_2, ϕ_2 and all configurations were independent. Easily modifying loop parameters in my python scripts and passing them as job array variables in a bash.sh linux script effectively parallelized the for loops that were submitted through a high-throughput qsub interactive submission system. I was also able to analyze results in real-time that were stored in my \h: home drive. The entire compute took about 3 days to complete.

Once, I began gathering results, I started constructing contour sequence plots like the one shown in figure 4.3. It became immediately apparent that optimal designs, when considering dual rotations for the off-diagonal $\mathcal{E}(\hat{\rho}_{ij})$ estimation in QPT, were optimal when the two axes were or close to orthogonal. There was little variation for asymmetric and noisy channels where axes were close to but not orthogonal. This finding was replicated by all the channels that I considered. I extracted peak performance points using a simple search algorithm that I wrote. It looked at other points in the neighbourhood of the optimal point and picked a region with the highest total specified number of points beyond a predefined threshold.

However, before I dive into a comparison between non-orthogonal and orthogonal optimal points, I need to highlight some very obvious (in hindsight) flaws in my approach so far and including Section 4.2.

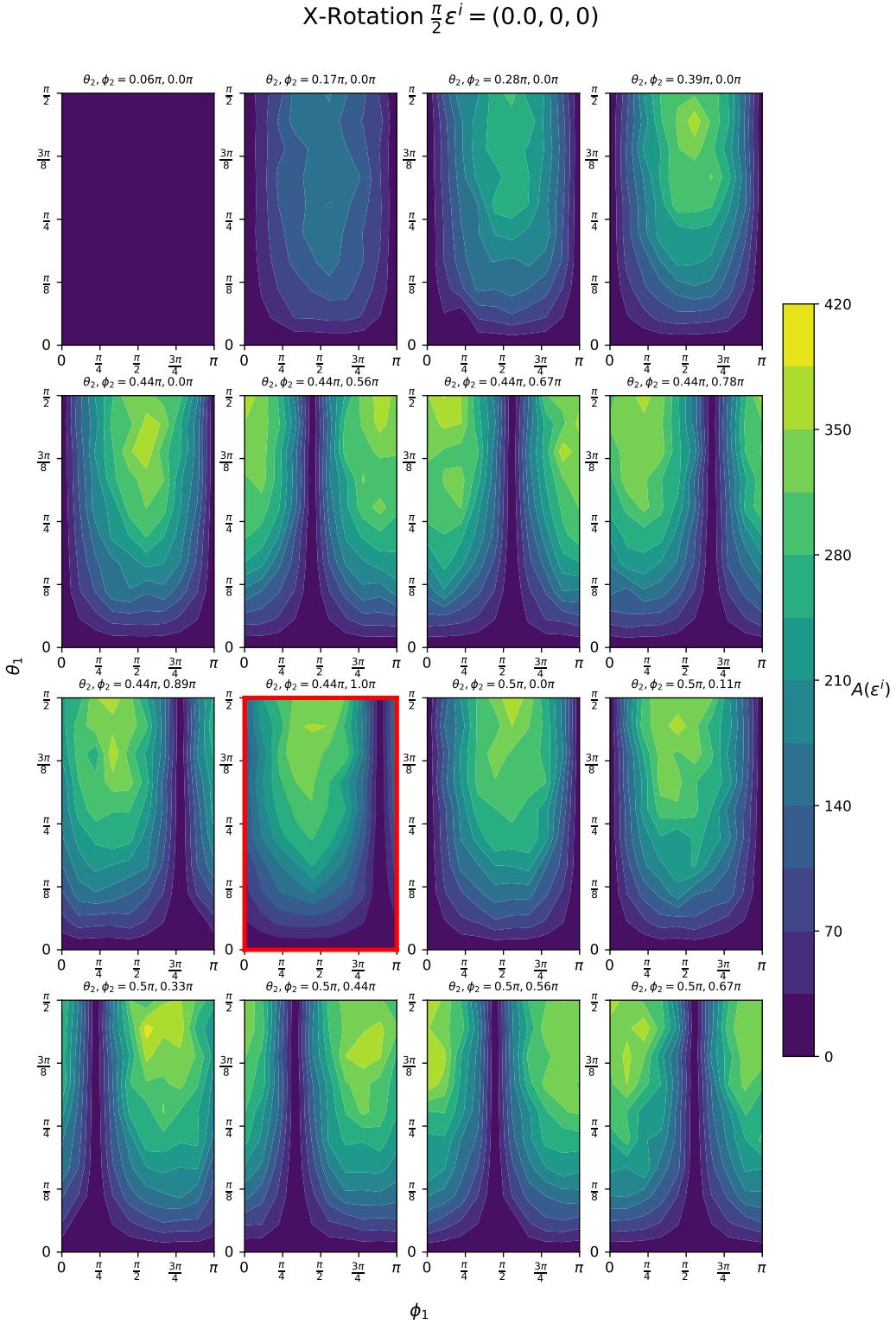


Figure 4.3: Contour sequence plot for varying x and y eigenstates in (4.6) for estimation of a $\frac{\pi}{2}$ unitary rotation about the x axis. The optimal scheme is highlighted in red and is used to normalize the color scale. Peaks start occurring when both rotating axes start approaching orthogonality. Poor performing landscapes can be seen in the first row. The middle and end rows are chosen configurations that were close to the orthogonal axes. Their abundance is an artefact of non-uniform polar sampling. In this particular case, the optimal points are close to orthogonal at $(\theta_1, \phi_1, \theta_2, \phi_2 = 0.44\pi, \pi, \frac{\pi}{2}, \frac{\pi}{2})$.

4.4 Issues with the rotations' approach

One very obvious flaw in the preceding approach is that the grid-sweeping angular steps might be too large. The resolution of the performance-angle surface, e.g. in figure 4.1b and especially figure 4.3, might not be sufficient to resolve optimal points, if they exist. However, the computational scheme forbade better resolutions.

Secondly, covering the Bloch sphere using polar angles does result in uniform covering. The angles close to the poles have a greater resolution than those closer to the equator. This, however, is a popular theoretical problem [47]. An astrophysics python library called HEALpix [48] that implements various spherically uniform sampling algorithms. Sadly, I did not have the time to pursue these avenues and it was also becoming increasingly likely that these approaches would not yield optimal preparation designs for the 1-qubit SQPT. Computational optimizations of this nature could be part of future endeavors if exploring projected subspaces of initial state d.o.f.s is ever necessary.

Another “possible” flaw is that the z-axis is fixed as the computational basis. This is an external asymmetry of the computational model that I was able to remove by the time I started gathering results for [Section 4.5](#). However, this implies that single axis rotations of the preparation design states were not all be identical and there might be a special price of fixing the computational basis that is not being accounted for in the above prescriptions.

Lastly, the primary flaw in the preceding approach is that the performance-angle surface is fluctuating, roughly in the range of 10 – 20%. The major source of those fluctuations is the inherent quantum noise due to finite ensemble measurement scheme used to compute $A(\epsilon^i)$ and then to a lesser extent, the variation in convergence of the constant noise model. This, I was able to easily remedy by iterating each point corresponding to a finite measurement ensemble (increasing in steps of 100 measurements as opposed to 10) on the power law decay curve (see figure 3.2) 100 times and computing the average performance factor $\tilde{A}(\epsilon^i)$ from the more stabilized average decay curve. This was an intensive computation and by the time I was able to realize this flaw, I had already submitted, received and completed the analysis of the datasets for [Section 4.3](#).

It turns out that the final flaw was very relevant to the non-orthogonal optimal points that I had found earlier. This became apparent when I implemented the iterations-averaging method for these non-trivial The results are plotted in figure 4.4 for all channels under consideration and are not surprising: none of the points that I had found were optimal!

4.5 Change of basis: selective exploration of preparation designs

Instead of trying to vary d.o.fs via brute-force which was computationally unfeasible over 2 2 d.o.fs, it made more sense to *guess* which designs would be most optimal and work backwards. The first design which, naturally, comes to mind is the

4 Results: varying preparation designs

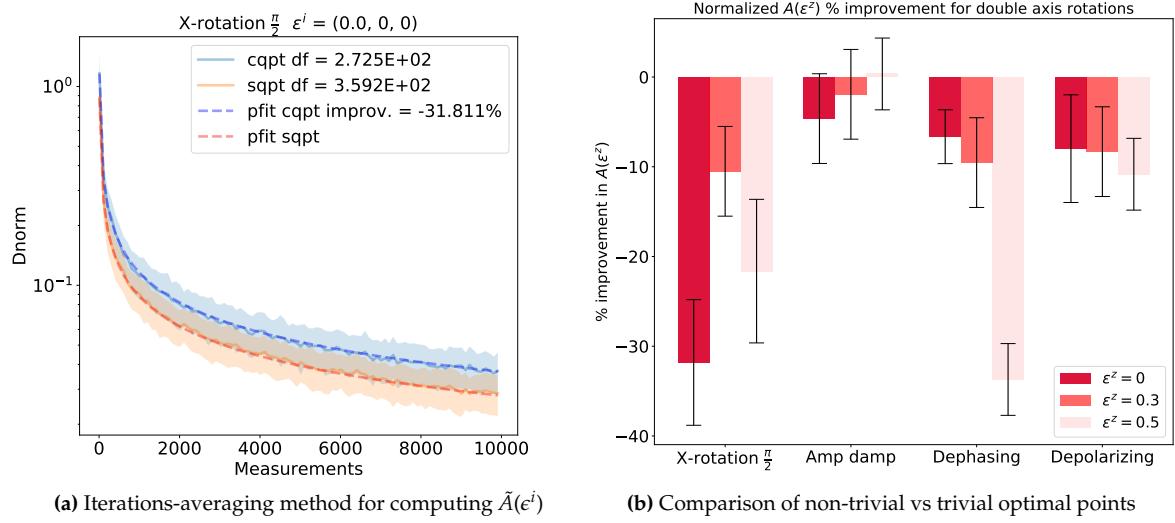


Figure 4.4: Applying iterations-averaging to accurately assess improvement in performance of the non-trivial optimal points of dual axis rotations reveals approximately no benefit. In a) this method is used with 100 iterations to compute average performance factors $\tilde{A}(\epsilon^i)$ for the SQPT (labelled “sqpt df”) and for the dual rotations (labelled “cqpt df”). A 2σ (standard deviation) of the iterations is also highlighted. The percentage changes between both designs are collected in b) along with the associated fluctuations for three noise levels $\epsilon^x = (0, 0.3, 0.5)$.

minimal informationally complete ² measurement set for 1-qubit: the tetrahedral set. Example Bloch vectors are,

$$\vec{r}_1 = (1, 1, 1) / \sqrt{3} \quad \vec{r}_2 = (1, -1, -1) / \sqrt{3} \quad \vec{r}_3 = (-1, 1, -1) / \sqrt{3} \quad \vec{r}_4 = (-1, -1, 1) / \sqrt{3} \quad (4.9)$$

or equivalently,

$$\{(\theta_i, \phi_i)\}_{tet} = \left\{ \left(\sin^{-1}\left(\frac{2}{3}\right), \frac{\pi}{4} \right), \left(\sin^{-1}\left(\frac{2}{3}\right), \frac{5\pi}{4} \right), \left(\pi - \sin^{-1}\left(\frac{2}{3}\right), \frac{3\pi}{4} \right), \left(\pi - \sin^{-1}\left(\frac{2}{3}\right), \frac{7\pi}{4} \right) \right\} \quad (4.10)$$

It is important to note that this set is not orthogonal,

$$\vec{r}_i \cdot \vec{r}_j = \frac{4}{3} \delta_{ij} - \frac{1}{3} \quad (4.11)$$

This crops up in the context of measurements: as a simple optimization to the number of observables needed to characterize a 1-qubit. Although it is common practice (like in my model) to work with the Pauli basis (4.1) which has 6 distinct eigenvalues for each operator (3 x 2 rank-1 projections), the minimal set on the other hand has 4 positive operators with each quantum state that can be physically prepared [26]. Now one can exploit the symmetry in the preparation and measurement d.o.f and consider using the tetrahedral set as a preparation design in (3.6).

There are two reasons why this choice is important:

1. It is *minimal*: the tetrahedral set of initial states is the smallest possible set of points on the Bloch sphere which are at equal distances from each other.

²this refers to the smallest number of linearly independent vectors needed to specify a d dimensional quantum state which is d^2 where $d = 2^n$ for the n -qubit. It is equivalent to the size of the span of the positive operator set, in terms of which, the quantum state as a density matrix can be decomposed. Recall that the span has 4 operators for the single qubit.

4 Results: varying preparation designs

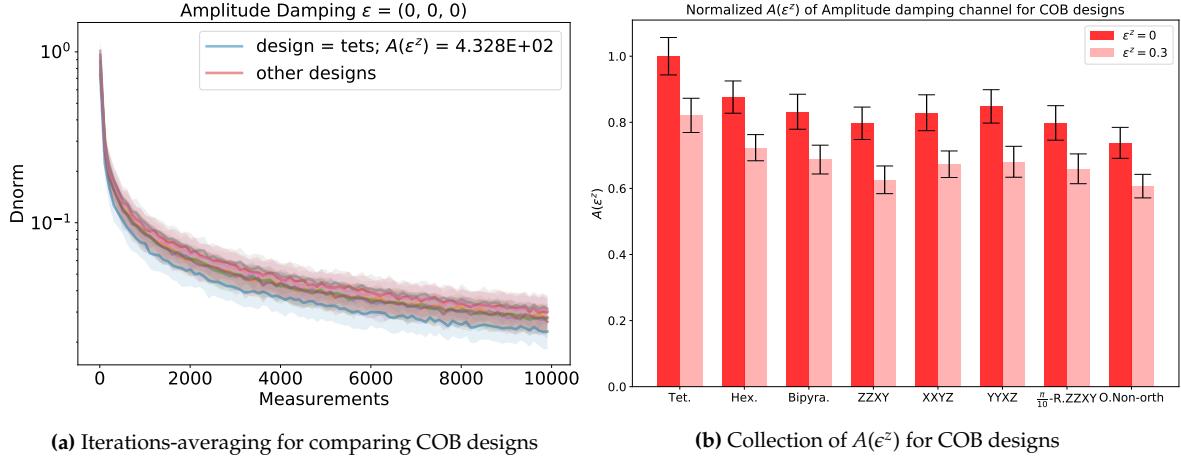


Figure 4.5: Applying iterations-averaging (with 100 iterations) to test proposed necessary conditions for optimality for various Change of Basis (COB) designs. Here, only the performance factors $A(\epsilon^z)$ for the amplitude damping channel $\mathcal{D}(\gamma = 0.2)$ are shown under a z-axis noise. The tetrahedral set outperforms all other designs, followed by the hexagonal set and then the bipyramedal set. In a) the specific decay curves are plotted and in b) the performance factors are collected for comparison and are normalized by the highest performing value. Orthogonal eigenstates are labelled by axes which when mentioned twice implicitly denote that the choice includes antipodal points on the axis.

2. It is *symmetric*: there are no special *off-diagonal* elements in this prescription that need to be physically and asymmetrically decomposed e.g. the standard choice (4.6)

Note that we are still working with the fixed Pauli measurement design. The tetrahedral design has the effect of evenly distributing XYZ eigenvalues in the estimators of all transformed channel states $\mathcal{E}(\hat{\rho}_{ij})$ needed for its characterization – bar any asymmetries associated with the channel action itself. This arguably causes better quantum and classical noise distribution that result from finite-size ensembles of different Pauli measurements in the QPT process.

Figure 4.5 shows a computational test for both the minimal and symmetry conditions put forward above. The conditions are computationally necessary for good performance. Using the iterations-averaging approach from the previous section, I compare performances of orthogonal, optimal non-orthogonal (I repeated estimation again due to the issue of fixing a basis in earlier approaches), symmetric but not minimal (bipyramedal, hexagonal) and the symmetric and minimal tetrahedral set. Note that the bipyramedal set has 5 points and the hexagonal set has six points corresponding to the edges of their 3-D representations on the Bloch sphere. The ensemble sizes per point in figure 4.5a were scaled appropriately to ensure that all designs received the same amount of measurements in total along the combined 3 axes.

As shown for the 1-qubit amplitude damping channel QPT in the noisy and non-noisy cases, similar results were replicated for all other channels in this study. They all show symmetric designs outperforming orthogonal and non-orthogonal designs. The tetrahedral design, satisfying both the minimal and symmetry conditions, outperforms all other designs and is an optimal design. Finally, the non-orthogonal optimal points found in Section 4.3 were again always outperformed by orthogonal designs.

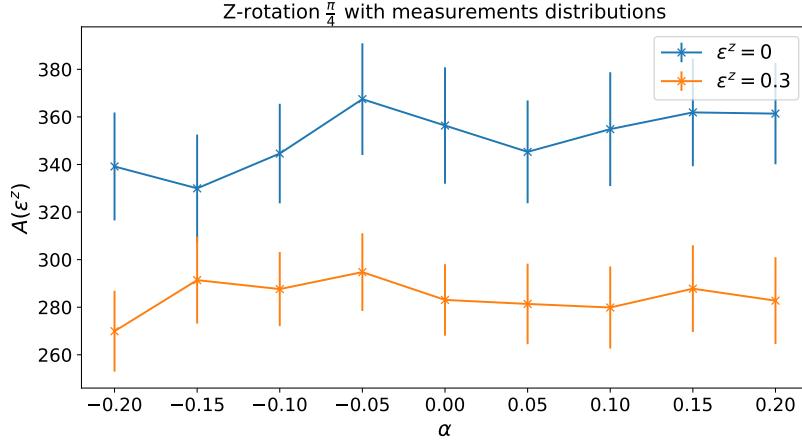


Figure 4.6: Effects of distributing number of measurements within an ensemble when the z-axis is noisy and not noisy ($\epsilon^z = 0, 0.3$) for a unitary rotation of $\frac{\pi}{4}$ are masked by quantum fluctuations. This is evidenced by the size of the error bars relative to changes in the distribution parameter α .

4.6 Optimizing performance within a preparation design

A natural question, when it comes to noisy preparation or measurement designs, is to ask whether we can redistribute a fixed total ensemble size (for states or measurement operators) within a certain design choice to optimize performance. For instance, if we are aware that the Z measurement axis is noisy, will increasing or decreasing the ensemble size of $\hat{\sigma}_z$ operators to lead to increases in $A(\epsilon^z)$ i.e. performance?

One approach to answer this would be to insert a displacement α in the number of measurements (N_x, N_y, N_z) for each axis within an ensemble, for each step-size, and repeat the power law decay fit with iterations-averaging. For the noisy z axis this will be,

$$(N_x, N_y, N_z) \rightarrow ((1 - \alpha)N_x, (1 - \alpha)N_y, (1 + 2\alpha)N_z)$$

where $\alpha \in [-0.3, 0.3] \in \mathbb{R}$ in steps of 0.05 and $N_x = N_y = N_z$ initially. Results are reported in figure 4.6 for a unitary rotation about channel about the x axis with z-axis noise $\epsilon^z = 0, 0.3$. It appears that the quantum fluctuations, for 100 iterations and averaging, are too large and mask any effects of the noise distribution, as evidenced by the size of fluctuations without the added noise. Histograms like figure 4.5b for unitary rotations and the symmetric depolarizing channel also revealed minute differences across noisy orthogonal designs.

Chapter 5

Conclusions

In this project, I have built a computational standard quantum process tomography (CQPT) protocol ([algorithm 1](#)) to quantitatively study the effects of varying preparation design stage of QPT while keeping the measurement design fixed. In [Section 3.3](#), a constant linear noise ϵ^i model was incorporated into CQPT for the aim of seeking robust designs and trade-offs in error with redistribution of measurements between M -ensembles axes' were studied. In [Section 3.5](#), a performance factor $A(\epsilon^i)$ model based on $O(1/\sqrt{M})$ power-law decay of a worst-case error (Dnorm), introduced in [Section 2.6](#), was used to make comparisions between different QPT preparation designs. Specifically, doing this for the single qubit, which requires at least 4 physical input states for QPT, produced two necessary conditions in [Section 4.4](#) for optimal designs: minimalness and symmetry, in terms of locations of the states on the Bloch sphere. Selective exploration of configurations produced the tetrahedral design that best satisfied these. Exploring variations of just 1 and 2 states in [Section 4.2](#) while keeping the others fixed revealed no optimization over the standard parametrization using Pauli eigenstates. Moreover, in [Section 4.6](#), it was found that the noise model was qualitative and weaker when compared with intrinsic quantum noise (due to probabilistic sampling) of the QPT problem as the performance did not respond to changing sizes of measurement-axis ensembles by redistributing the total allowed measurements in the presence of axis noise.

In terms of the aims of this thesis, the second aim was most successfully tackled. Two necessary conditions for optimality of preparation designs were discovered based on arguments of symmetrizing QPT and evenly distributing asymmetries. However, due to a lack of time and expertise, this argument could have been made more mathematical and quantitative using ideas from probability theory. Based on the conditions, I conjecture that the tetrahedron is optimal for larger qubits and the argument might generalize to hyper tetrahedrons for n -qubit bloch hyperspheres as the number of points are the same as the basis vector elements needed to specify a density operator or a quantum state. The hypertetrahedron vertices would be maximally seperated on the surface of the bloch hypersphere. However, the entire surface of hyperspheres is not physically parametrizable as quantum states so this might not be the case. A proper argument using stricter framework is needed to solidify this hunch.

The third aim, which was to simulate realistic QPT noise, had marginal success. The model that I decided to work with had some nice properties: it converged in the

limit of large measurements; it vertically translated Dnorm decay curves which had a proportionate affect on performance factors. It was also linear and did not exotically affect Dnorm decay with non-trivial measurement statistics, which was crucial to do quantitative variation of preparation designs. However, the noise converged too quickly which was possibly why everything else was so lightly affected. As there were two competing sources of noise in the QPT protocol, quantum and classical, it was not obvious how to calibrate the two sources of noise so that they could be comparable enough to simulate real experiments. To go about optimizing the noise model, one place to start would be the measurements redistribution problem under single axis noise in [Section 4.6](#). Inspiration could be taken from real experiments, something I did not have the time to study in detail. Other forms of noise like Poisson noise from quantum optics and fancier methods like convolutions with noise and then applying deconvolving corrections e.g. finding the biorthogonal bases method for noisy observables could be tried [49]. However, a better avenue in my opinion, is the creation or, if it already exists, integration of a probability distribution with all the desired properties discussed above to be used as a source to sample e^i . The effects of theoretical convolutions with multinomial measurement outcomes could be understood better as a result as well. An alternative would be to explore theoretically optimal Bayesian adaptive strategies for noise as well as QST [50].

Lastly, the first aim was to find precise effects of preparation design on QPT. This was only indirectly touched upon in this thesis. The vision behind this aim was to understand preparation design variation to a high degree to be able to propose a new QPT scheme. This, however, was not achieved in the time-frame that I was working with if it was indeed even possible. If this wasn't clear, I would have worked on establishing constraints on the existence of QPT schemes around optimizations of preparation designs. The tetrahedral design performance and the optimality conditions were only discovered near the end of the project. One can potentially work backwards from this to create a direct and online QPT method centered on using optimal designs. A different paradigm of *weak* measurements that don't collapse quantum states like those proposed in [41] is also another avenue to start with that is different and might mesh well with newer optimal state designs proposed in this thesis.

Bibliography

- [1] Christopher Ferrie. Some theory and applications of probability in quantum mechanics. 2012. [doi:10.1101/6822](https://doi.org/10.1101/6822).
- [2] John S Bell. On the einstein podolosky rosen paradox, physics. *Reprinted in Quantum Theory and Measurement*, 1:195–200, 1964. [doi:10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195).
- [3] Marissa Giustina, Marijn AM Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, et al. Significant-loophole-free test of bell's theorem with entangled photons. *Physical review letters*, 115(25):250401, 2015. [doi:10.1103/PhysRevLett.115.250401](https://doi.org/10.1103/PhysRevLett.115.250401).
- [4] Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. Strong loophole-free test of local realism. *Physical review letters*, 115(25):250402, 2015. [doi:10.1103/PhysRevLett.115.250402](https://doi.org/10.1103/PhysRevLett.115.250402).
- [5] John Von Neumann. *Mathematical foundations of quantum mechanics: New edition*. Princeton university press, 2018.
- [6] Michael A. Nielsen and Isaac Chuang. Quantum computation and quantum information. *American Journal of Physics*, 70(5):558–559, 2002. [doi:10.1119/1.1463744](https://doi.org/10.1119/1.1463744).
- [7] Hartmut Häffner, Wolfgang Hänsel, CF Roos, Jan Benhelm, Michael Chwalla, Timo Körber, UD Rapol, Mark Riebe, PO Schmidt, Christoph Becher, et al. Scalable multiparticle entanglement of trapped ions. *Nature*, 438(7068):643–646, 2005. [doi:10.1038/nature04279](https://doi.org/10.1038/nature04279).
- [8] Wei-Bo Gao, Chao-Yang Lu, Xing-Can Yao, Ping Xu, Otfried Gühne, Alexander Goebel, Yu-Ao Chen, Cheng-Zhi Peng, Zeng-Bing Chen, and Jian-Wei Pan. Experimental demonstration of a hyper-entangled ten-qubit schrödinger cat state. *Nature physics*, 6(5):331–335, 2010. [doi:10.1038/nphys1603](https://doi.org/10.1038/nphys1603).
- [9] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355, 1998. [doi:10.1103/PhysRevE.58.5355](https://doi.org/10.1103/PhysRevE.58.5355).

- [10] Antonio Acín, Immanuel Bloch, Harry Buhrman, Tommaso Calarco, Christopher Eichler, Jens Eisert, Daniel Esteve, Nicolas Gisin, Steffen J Glaser, Fedor Jelezko, et al. The quantum technologies roadmap: a european community view. *New Journal of Physics*, 20(8):080201, 2018. [doi:10.1088/1367-2630/aad1ea](https://doi.org/10.1088/1367-2630/aad1ea).
- [11] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear algebra and its applications*, 10(3):285–290, 1975. [doi:10.1016/0024-3795\(75\)90075-0](https://doi.org/10.1016/0024-3795(75)90075-0).
- [12] Andrzej Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972. [doi:10.1016/0034-4877\(72\)90011-0](https://doi.org/10.1016/0034-4877(72)90011-0).
- [13] Yazhen Wang et al. Asymptotic equivalence of quantum state tomography and noisy matrix completion. *The Annals of Statistics*, 41(5):2462–2504, 2013. [doi:10.1214/13-AOS1156](https://doi.org/10.1214/13-AOS1156).
- [14] Masoud Mohseni, AT Rezakhani, and DA Lidar. Quantum-process tomography: Resource analysis of different strategies. *Physical Review A*, 77(3):032322, 2008. [doi:10.1103/PhysRevA.77.032322](https://doi.org/10.1103/PhysRevA.77.032322).
- [15] Zhi-Wei Wang, Yong-Sheng Zhang, Yun-Feng Huang, Xi-Feng Ren, and Guang-Can Guo. Experimental realization of direct characterization of quantum dynamics. *Physical Review A*, 75(4):044304, 2007. [doi:10.1103/PhysRevA.75.044304](https://doi.org/10.1103/PhysRevA.75.044304).
- [16] Masoud Mohseni and DA Lidar. Direct characterization of quantum dynamics. *Physical review letters*, 97(17):170501, 2006. [doi:10.1103/PhysRevLett.97.170501](https://doi.org/10.1103/PhysRevLett.97.170501).
- [17] Ariel Bendersky, Fernando Pastawski, and Juan Pablo Paz. Selective and efficient quantum process tomography. *Physical Review A*, 80(3):032116, 2009. [doi:10.1103/PhysRevA.80.032116](https://doi.org/10.1103/PhysRevA.80.032116).
- [18] Christoph Dankert. Efficient simulation of random quantum states and operators, 2005. [arXiv:quant-ph/0512217v2](https://arxiv.org/abs/quant-ph/0512217v2).
- [19] Isaac L Chuang and Michael A Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997. [doi:10.1080/09500349708231894](https://doi.org/10.1080/09500349708231894).
- [20] Christopher Ferrie. Self-guided quantum tomography. *Physical review letters*, 113(19):190404, 2014. [doi:10.1103/PhysRevLett.113.190404](https://doi.org/10.1103/PhysRevLett.113.190404).
- [21] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014. [doi:10.1038/nphys3029](https://doi.org/10.1038/nphys3029).
- [22] Alexander I Lvovsky and Michael G Raymer. Continuous-variable optical quantum-state tomography. *Reviews of modern physics*, 81(1):299, 2009. [doi:10.1103/RevModPhys.81.299](https://doi.org/10.1103/RevModPhys.81.299).

5 BIBLIOGRAPHY

- [23] A Shabani, RL Kosut, M Mohseni, H Rabitz, MA Broome, MP Almeida, A Fedrizzi, and AG White. Efficient measurement of quantum dynamics via compressive sensing. *Physical review letters*, 106(10):100401, 2011. [doi:10.1103/PhysRevLett.106.100401](https://doi.org/10.1103/PhysRevLett.106.100401).
- [24] Christopher J. Wood. Non-completely positive maps: properties and applications, 2009. [arXiv:0911.3199](https://arxiv.org/abs/0911.3199).
- [25] John Watrous. *The theory of quantum information*. Cambridge University Press, 2018. [doi:10.1017/9781316848142](https://doi.org/10.1017/9781316848142).
- [26] Yong Siah Teo. *Introduction to quantum-state estimation*. World Scientific, 2016. [doi:10.1142/9789814678858_0002](https://doi.org/10.1142/9789814678858_0002).
- [27] Edwin KP Chong and Stanislaw H Zak. *An introduction to optimization. Fourth edition*. John Wiley & Sons, 2008. [doi:10.1002/9781118033340](https://doi.org/10.1002/9781118033340).
- [28] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2012. [doi:10.1007/978-1-4615-4381-7](https://doi.org/10.1007/978-1-4615-4381-7).
- [29] Aharon Brodutch, Animesh Datta, Kavan Modi, Ángel Rivas, and César A Rodríguez-Rosario. Vanishing quantum discord is not necessary for completely positive maps. *Physical Review A*, 87(4):042301, 2013. [doi:10.1103/PhysRevA.87.042301](https://doi.org/10.1103/PhysRevA.87.042301).
- [30] Bassano Vacchini and Giulio Amato. Reduced dynamical maps in the presence of initial correlations. *Scientific reports*, 6(1):1–12, 2016. [doi:10.1038/srep37328](https://doi.org/10.1038/srep37328).
- [31] GM D’Ariano and P Lo Presti. Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. *Physical review letters*, 86(19):4195, 2001. [doi:10.1103/PhysRevLett.86.4195](https://doi.org/10.1103/PhysRevLett.86.4195).
- [32] Benjamin Schumacher. Sending entanglement through noisy quantum channels. *Physical Review A*, 54(4):2614, 1996. [doi:10.1103/PhysRevA.54.2614](https://doi.org/10.1103/PhysRevA.54.2614).
- [33] Michael A Nielsen. The entanglement fidelity and quantum error correction. *arXiv preprint quant-ph/9606012*, 1996. [arXiv:quant-ph/9606012](https://arxiv.org/abs/quant-ph/9606012).
- [34] Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states. *Proceedings of the thirtieth annual ACM symposium on Theory of computing - STOC ’98*, 1998. [doi:10.1145/276698.276708](https://doi.org/10.1145/276698.276708).
- [35] John Watrous. Simpler semidefinite programs for completely bounded norms, 2012. [arXiv:1207.5726](https://arxiv.org/abs/1207.5726).
- [36] Avraham Ben-Aroya and Amnon Ta-Shma. On the complexity of approximating the diamond norm, 2009. [arXiv:0902.3397](https://arxiv.org/abs/0902.3397).
- [37] Larry Wasserman. Lecture 8: Minimax theory. URL: <http://www.stat.cmu.edu/~larry/=stat705/Lecture8.pdf>.

5 BIBLIOGRAPHY

- [38] Mário Ziman, Martin Plesch, Vladimír Bužek, and Peter Štelmachovič. Process reconstruction: From unphysical to physical maps via maximum likelihood. *Physical Review A*, 72(2):022106, 2005. [doi:10.1103/PhysRevA.72.022106](https://doi.org/10.1103/PhysRevA.72.022106).
- [39] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017. [doi:10.5555/1211332](https://doi.org/10.5555/1211332).
- [40] Akio Fujiwara and Paul Algoet. One-to-one parametrization of quantum channels. *Physical Review A*, 59(5):3290, 1999. [doi:10.1103/PhysRevA.59.3290](https://doi.org/10.1103/PhysRevA.59.3290).
- [41] George C Knee. *Concepts and applications of quantum measurement*. PhD thesis, Oxford University, UK, 2014. URL: <https://ora.ox.ac.uk/objects/uuid:2838a30b-302c-4fac-9e86-1ca452a88a83>.
- [42] Jaroslav Řeháček, Zdeněk Hradil, E Knill, and AI Lvovsky. Diluted maximum-likelihood algorithm for quantum tomography. *Physical Review A*, 75(4):042108, 2007. [doi:10.1103/PhysRevA.75.042108](https://doi.org/10.1103/PhysRevA.75.042108).
- [43] George C Knee, Eliot Bolduc, Jonathan Leach, and Erik M Gauger. Quantum process tomography via completely positive and trace-preserving projection. *Physical Review A*, 98(6):062336, 2018. [doi:10.1103/PhysRevA.98.062336](https://doi.org/10.1103/PhysRevA.98.062336).
- [44] Anastasios Kyrillidis, Amir Kalev, Dohyung Park, Srinadh Bhojanapalli, Constantine Caramanis, and Sujay Sanghavi. Provable compressed sensing quantum state tomography via non-convex methods. *npj Quantum Information*, 4(1):1–7, 2018. [doi:10.1038/s41534-018-0080-4](https://doi.org/10.1038/s41534-018-0080-4).
- [45] Christian Schwemmer, Lukas Knips, Daniel Richart, Harald Weinfurter, Tobias Moroder, Matthias Kleinmann, and Otfried Gühne. Systematic errors in current quantum state tomography tools. *Physical review letters*, 114(8):080403, 2015. [doi:10.1103/PhysRevLett.114.080403](https://doi.org/10.1103/PhysRevLett.114.080403).
- [46] Imperial college research computing service. 2020. [doi:10.14469/hpc/2232](https://doi.org/10.14469/hpc/2232).
- [47] Anna Yershova, Swati Jain, Steven M Lavalle, and Julie C Mitchell. Generating uniform incremental grids on so (3) using the hopf fibration. *The International journal of robotics research*, 29(7):801–812, 2010. [doi:10.1177/0278364909352700](https://doi.org/10.1177/0278364909352700).
- [48] Krzysztof M Gorski, Eric Hivon, Anthony J Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthias Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, 2005. [doi:10.1086/427976](https://doi.org/10.1086/427976).
- [49] Giacomo Mauro D’Ariano. Universal quantum estimation. *Physics Letters A*, 268(3):151–157, 2000. [doi:10.1016/S0375-9601\(00\)00164-X](https://doi.org/10.1016/S0375-9601(00)00164-X).
- [50] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, 2010. [doi:10.1088/1367-2630/12/4/043034](https://doi.org/10.1088/1367-2630/12/4/043034).