Verification and Validation of Quantum Systems

By

SAM PALLISTER

Quantum Engineering Centre for Doctoral Training
and
School of Mathematics

UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

OCTOBER, 2018

Word count: Forty-four thousand.

ABSTRACT

The goal of this thesis is to explore the extent and limitations of existing techniques for the verification and validation of quantum systems, and to develop new techniques that are: (i) more efficient; (ii) mathematically rigorous; and (iii) practically implementable.

We give a survey of contemporary techniques and protocols for the verification and characterisation of quantum states and processes, before deriving efficient protocols of our own. We construct an optimally efficient protocol constructed from local measurements for verifying that the output of an experiment agrees with a known target state, for both two-qubit and stabilizer states. We then show that this protocol is also optimal for the task of estimating the fidelity to a known target state, and to analyse its effectiveness we run both the optimal protocol and its closest competitors on a silicon-based photonic chip designed to produce arbitrary two-qubit states.

We discuss the validation of quantum systems in two contexts. Firstly, we validate a quantum computational advantage over classical computers when faced with the task of producing solutions to differential equations via the finite element method. Secondly, we provide a blueprint for the validation of quantum effects in general relativistic systems, by outlining a space-based experiment designed to test the interplay between the two theories.

ACKNOWLEDGEMENTS

I would first like to thank my supervisory singlet, Ashley Montanaro and Noah Linden, for being coherent, correlated, and only marginally mixed; and for being a powerful resource for both error correction and quantum steering.

I must also thank my friends and peers, whose support was invaluable to my sanity over the last four years: Chris Cade and Stephen Piddock, for the sofas and cakes; Euan Allen, Matt Day and Sam Morley-Short, for the beers and cheese.

I'd also like to express my gratitude to colleagues, collaborators and friends in the CDT; and to the CDT management team themselves, for affording me the opportunity to act as a lab rat in their madcap experiment, and for sufficiently baiting the maze to get me out the other end.

And finally, my greatest love and thanks are reserved for Jen, whose affection and companionship I am eternally grateful for.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

This thesis is dedicated to my parents:
to my Dad, for getting me here in style; and
to my Mum, for getting me here in one piece.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Quantum systems

This dissertation concerns *quantum systems*; that is, systems that fundamentally require quantum mechanics to describe their operation. Of particular importance are *quantum technologies*; technologies that process information in a fundamentally quantum mechanical way, to achieve a particular goal. One might be tempted to include venerable 20^{th} century technologies such as the transistor or the laser in this categorisation, as the most faithful physical model of their operation is a quantum mechanical one. However, the information processing of a transistor is sufficiently described by classical information theory, and its operation could be mimicked (albeit with a probable cost in both size and ease of use) by a fully classical device. Instead, we will be concerned with devices that use quantum mechanical effects, such as *superposition* and *entanglement*, to process information quantum mechanically. In particular, we have the following emerging technologies in mind:

1. **Quantum computing**.
2. **Quantum communications and cryptography**.
3. **Quantum sensing and imaging**.

Theoretical and experimental progress on each of these technologies is being made at an ever accelerating rate. It is already possible to buy commercial quantum cryptography systems, and devices have been developed that are compact and robust enough to be space-qualified and operated in orbit [88]. Key components of quantum sensing systems, such as atomic clocks and frequency combs, are also commercially available. In quantum computing, the size and complexity of hardware is developing rapidly after a period of slow growth, with the number of controllable qubits increasing by an order of magnitude in the last two years [73, 168, 126]. This is

progressing in tandem with the development of both more elaborate and practical pieces of quantum software and quantum algorithms [161, 123]. As the size and complexity of these devices increases, development of suites of verification tools becomes imperative; both for the understanding of the current iteration of quantum technologies, and to inform the design and operation of future iterations.

1.2 Verification, validation and characterisation

The tasks of verification, validation and characterisation are both computationally and intellectually intensive for any complex system, not just for those described by quantum mechanics. Quantum systems are not singular in having rich and complicated mathematical structure, nor in having a large and rapidly expanding number of highly engineered components that need to be tested to specification. These testing procedures are critical to the successful operation of most complex devices. As a case in point, on the 11^{th} of December, 1998, NASA launched the *Mars Climate Orbiter*, a $327m mission to explore the changing climate and water composition of Mars over geologic time scales. The orbiter would travel almost 700 million kilometres over the course of the next 300 days, culminating in an interception with Mars and an insertion in its orbit. On the 23^{rd} of September, 1999, this orbital insertion was initiated; the process began at around 08:45. By 09:15, the orbiter had completely disintegrated, vaporised by the drag caused by the Martian atmosphere. The cause was a miscalibration of the trajectory of the orbiter. Over the course of its nine month journey the orbiter, like most interplanetary missions, needed to calibrate and correct its path by firing small thrusters on the side of craft. A piece of on board software would calculate the forces and torques on the orbiter from the thruster firings, broadcast the data to Earth, and a computation on Earth would revise the predicted trajectory. However, the on board software, written by Lockheed Martin, calculated the forces and torques in imperial units before broadcasting them to Earth. NASA's software, expecting metric units, erroneously corrected for a force that was twice the magnitude of the true force, culminating in a trajectory that deviated enough to send the orbiter deep into Mars' atmosphere. While the error was avoidable, the failure of systems to verify the calculation was inexcusable. NASA's official incident report cited failure of verification and validation systems as a key cause of the mission's failure [134]. If we are to successfully operate quantum devices that are of a comparable size and complexity, we must heavily scrutinise our ability to test them effectively.

The nomenclature "verification, validation and characterisation" often appears as a triplet, and colloquially, we often use each term semi-synonymously. However,

in most contexts (and particularly in the quantum context that will form the backbone of this thesis), they are not interchangeable; they satisfy distinct roles in the assessment of systems. We will use the following operational definitions [189]:

- **Verification:** The process by which a system is tested to determine whether it satisfies a particular specification. The specification need not be a fully parameterised model of the system; it may be a certain set of requirements, properties or criteria.
- **Characterisation:** An instance of a verification task, where the specification explicitly *is* a fully parameterised model of the system.
- **Validation:** The process by which the specification itself is scrutinised, to determine whether it is stringent enough for the device or system to successfully complete a task on behalf of a user.

We will shortly discuss each of these tasks in detail in the context of quantum systems. In the context of verification of classical software and hardware, some important distinctions are made between different families of verification protocols. The first distinction to be made is between *formal verification* methods, and *empirical verification* methods. Formal verification concerns guaranteeing the correctness of a system by strict mathematical or logical deduction. These methods typically require three steps: (i) the construction of a suitable mathematical model of the system being interrogated; (ii) a specification that outlines a set of salient parameters from the model and acceptable ranges of values; and (iii) some logical or mathematical machinery that maps from the model to a statement about whether the system meets the specification. This proof machinery can either be worked through by hand, semi-automated in the sense that the proof is provided by a user and checked by computer, or fully delegated to a computer that makes use of automated proof methods.

Furthermore, formal verification broadly breaks down into two subcategories: *deductive verification* and *model checking*. Model checking takes a comprehensive approach - given a specification of the system, model checking methods exhaustively explore the entire state space of the model to determine whether the system is to specification. In this sense, model checking is very closely related to the process of characterisation. Deductive verification refers to the construction of a set of proof obligations for the model and the specification, the correctness of which is sufficient to ensure that the system is to specification. Both of these protocols have utility. Model checking is, in principle, fully autonomous and requires minimal input from a verifier, and in broad classes of instances can be guaranteed to terminate with a consistent answer to the verification protocol, given the necessary computational

resources. Conversely, the computational resources that it requires blows up significantly with the size of the state space of the model, and so model checking very complex systems is computationally prohibitive [60]. This blow up in resources is combated (but not wholly defeated) by finding efficient representations of the model, or by exploiting some symmetry or modularity in the system [59]. Deductive verification, on the other hand, does not experience this state space explosion problem as acutely, but generally requires a large degree of human input to derive a set of sufficient proof rules from the specification that the system must be checked against.

In contrast to formal verification, *empirical verification* involves preparing a set of experiments for the model in the hope that this set captures the typical behaviour of the system, possibly coupled with some reasoning about the likelihood and mitigation of atypical cases. Empirical verification also subdivides into two categories: *simulation* and *testing*. Simulation refers to the scenario where a set of experiments are run on the model, whereas testing consists of running a set of experiments on the system directly. Both empirical verification methods can be run cheaply compared to their formal counterparts, but lack the strong mathematical guarantees that formal methods provide.

We will use this taxonomy to classify verification methods for quantum systems in Chapter 2.

1.2.1 Validation of quantum systems

As opposed to *verification*, the *validation* of quantum systems is an entirely different process. To be clear, validation is the process by which the *specification itself* is scrutinised, to determine whether it is stringent enough for the device or system to successfully complete a task on a user's behalf. Take the following quantum mechanical example: suppose a user wishes to generate, say, high fidelity Bell states in order to carry out a quantum cryptography protocol. Verification is the process of scrutinising the Bell states - is there an infidelity threshold at which the protocol fails? If so, are the Bell states of high enough fidelity for the protocol to succeed? Validation, on the other hand, is the process of scrutinising the protocol - does this cryptography protocol have the security guarantees that the user requires? Does it satisfy the user's needs with respect to key rate, or with respect to cost?

In most quantum technologies, the proposed benefit to the user is that there is some theoretically promised quantum advantage over the technology's classical counterparts. Hence the validation of quantum technologies requires showing that, given some reasonable specification of a quantum device, that it has some demonstrable quantum advantage that is of benefit to a hypothetical user. In this

sense, large subfields within quantum information fall under the umbrella of validation of quantum systems. For example, one of the main goals of quantum algorithms research is to demonstrate a quantum computational advantage where either the specification is of an ideal, noiseless quantum computer, or some representation of a physically realisable quantum computer. The same could be said of security proofs in quantum cryptography, or of precision bounds in quantum metrology. As a particular case study, determining the correctness of the quantum algorithm for solving differential equations presented in Chapter 5 can reasonably be argued to fall under the remit of validation of quantum systems.

While this notion of validation covers large subfields within quantum information, there is a tangential notion of validation that is worth exploring for quantum systems. For most quantum technologies, we assume an axiomatic, quantum mechanical framework that in principle is sufficient to describe the operation of that particular technology. However, it may be the case that such systems fail to behave non-classically, or that this framework is not sufficient in the setting in which the system operates (for example, a quantum cryptography satellite operating in a non-negligible gravitational potential). In these cases, "validation" can be construed as the examination of whether the system actually behaves quantum mechanically. We will return to this notion of validation in Chapter 6.

1.3 Thesis overview

The first half of this thesis concerns verification of quantum systems. In Chapter 2, we give an extensive literature review on contemporary techniques for verifying quantum states and processes. The goal is to use this as a foothold to develop new techniques for verifying quantum states in Chapters 3 and 4. In Chapter 3, we examine the verification of quantum states. We devise a framework for state verification protocols, and a set of desiderata that practically-implementable verification protocols should satisfy. Within this framework, we develop copy-optimal verification protocols for two-qubit states and stabilizer states that are quantitatively, and significantly, more efficient than the prior art. We also introduce an ansatz for a verification protocol that attains the same efficiency advantage, but is valid for verification of arbitrary pure states. In Chapter 4, we shift focus to fidelity estimation protocols. We again develop a framework and desiderata for protocols for this task, before deriving the most efficient estimator given those desiderata. We then apply this protocol on output data from a silicon photonic chip designed to produce arbitrary two-qubit states, and we compare the protocol's performance to common alternatives.

The second half of this thesis concerns validation of quantum systems. In Chapter 5, we give a quantum algorithm for solving differential equations via the finite element method, and give the first full analysis of the computational complexity of a quantum algorithm for this problem. We validate the existence of an explicit computational advantage over classical counterpart algorithms, and also give arguments as to why this advantage is unlikely to be significantly improved. In Chapter 6, we give a blueprint for a space-borne experiment designed to validate quantum effects in a general relativistic system. We give a full feasibility analysis of the experiment, including precise calculations of the effect we wish to test, a full description of the experimental payload, and an analysis of the experiment's feasibility given the current capabilities of space-qualified optical components.

1.4 Declaration of contributions and previous publications

The survey on the verification of quantum systems in Chapter 2 was written exclusively by the author.

Chapter 3, on the verification of quantum states, is based on research carried out with Noah Linden and Ashley Montanaro; the results on verification of two-qubit and stabilizer states were published in *Physical Review Letters* in March 2018 [178]. Both the premise and the list of protocol requirements in § 3.2.1 were developed jointly between the author, AM and NL. The calculation of optimal strategies for Bell states, two qubit states and stabilizer states in § 3.3, § 3.4 and § 3.5, respectively, were carried out by the author, with some important proof techniques (for example, the "strategy averaging" step in § 3.4) contributed by AM and NL. The extension to "soft" verification in § 3.6 is due to the author. Regarding the extensions to verification of arbitrary states in § 3.7, the statement of the ansatz in § 3.7.4 is due to AM and NL, but the remaining analysis is due to the author.

Chapter 4, on fidelity estimation on a photonic chip, is joint work with Xiaogang Qiang. Specifically, the entirety of the mathematical contributions are the author's, with the description and operation of the chip due to XQ (paraphrased from [190]).

Chapter 5, on quantum algorithms for solving partial differential equations, represents joint work with Ashley Montanaro. The content of this chapter was published in *Physical Review A* in March 2016 [164]. The analysis of the finite element method (FEM) in § 5.1 and § 5.3, as well as the classical complexity of solving FEM instances § 5.4, is due to the author. The framework for the quantum algorithm for the FEM in § 5.6 is due to AM, with the complexity analysis due to AM and the author. The limitations on the quantum algorithm in § 5.8 are due to AM.

Lastly, Chapter 6, on a blueprint for a simultaneous test of quantum mechanics and general relativity, was a collaboration between the author and a team of European colleagues on behalf of the European Space Agency. The content of this chapter was published in *European Physical Journal: Quantum Technology* in February 2017 [179]. The premise and original experimental proposal are due to the author. The calculations in the scientific background, § 6.2, are due to both the author and collaborators. The precise details of the payload in § 6.3 were devised by experimentalists on the team, and those of the mission design in § 6.5 by engineers on the team. Both aspects were organised and collated by the author. The Monte Carlo statistical analysis of the mission in § 6.4 is due to the author.

1.5 Nomenclature, conventions and definitions

We will make use of both conventions $\{\mathbb{1}, X, Y, Z\}$ and σ_k, $k = 0,1,2,3$ (where the index ordering matches the former convention) to denote the Pauli matrices. For tensor products of Pauli matrices, we will also use the symbol σ_k, but the dimensionality of the tensor product will be apparent from the range of the index k. Also, we will often use the shorthand $MN = M \otimes N$ (i.e. that the composition of symbols MN refers to the tensor product, rather than the matrix product). In general the type of product will be obvious from context but will be explicitly pointed out when necessary.

We use the "squared" definition for the fidelity between two states:

$$F(\rho,\sigma) \coloneqq \left[\mathrm{tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)\right]^2. \tag{1.1}$$

Unless explicitly stated elsewhere, in the context of verification the parameter ϵ will denote the infidelity between two states, $F(\rho,\sigma) = 1-\epsilon$. The trace distance is given by

$$T(\rho,\sigma) \coloneqq \frac{1}{2}\|\rho-\sigma\|_1 = \frac{1}{2}\mathrm{tr}\left[\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\right], \tag{1.2}$$

and in the context of verification we will often denote the trace distance with parameter ϵ_T. The infidelity and trace distance are related by the following inequalities:

$$1-\sqrt{1-\epsilon} \leq \epsilon_T \leq \sqrt{\epsilon}. \tag{1.3}$$

In the discussion on verification of quantum states in Chapters 2, 3 and 4, lower case n refers to the number of copies of a particular state ρ that must be consumed to carry out a particular task (the "copy complexity"). Upper case N refers to the number of qubits over which a state is defined; $|\psi\rangle \in \mathbb{C}^{2^N}$. We will alternatively use $d = 2^N$ when this is clearer in terms of notation. We denote the space of density matrices over $\mathbb{C}^d$ as $\mathcal{D}(\mathbb{C}^d)$.

CHAPTER 2

VERIFICATION AND CHARACTERISATION OF QUANTUM SYSTEMS: A SURVEY

When considering the verification of quantum systems, a plethora of techniques have been developed over the last 20 years with overlapping sets of advantages, drawbacks and domains of applicability. Also, the term "quantum system" is incredibly broad. As such we must be selective about the types of system and techniques that we will discuss in this chapter. Our focus will be on the verification and characterisation of quantum states and quantum processes. For systems describing quantum computations, an extensive and diverse body of literature regarding verifiability of their outputs (in particular, in the context of interactive proof systems) has been developed in recent years [86, 157, 35, 109, 87, 83, 151]. However, a full discussion of such systems is outside the scope of this survey.

Our goal will be to provide sufficient background in the field to introduce the results on state verification in Chapter 3, and the results on fidelity estimation in Chapter 4. The protocols developed in these chapters are given with a precise mathematical guarantee on their performance, and we claim that the performance of these protocols is advantageous with respect to the most closely-applicable prior art. We therefore consider it fruitful to first discuss the most common prior protocols for state verification, including: (i) what the protocol tests; (ii) its operational assumptions; (iii) the existence and quality of performance guarantees; and (iv) its relationship to the protocols in Chapters 3 and 4.

For both verification of quantum states and of quantum processes, the most prominent protocols and tools fit quite neatly into the taxonomy that has been long established for the verification of classical software and hardware [223, 58]; see Table 2.1. We will now give an overview of established techniques for verification and characterisation of quantum systems.

	Formal verification		**Empirical verification**	
Mantra	*Characterise entire model*	*Check sufficient criteria for model*	*Check typical model instances*	
Classical methods	Model checking	Deductive verification	Simulation	Testing
Verifying quantum states	Quantum state tomography (§ 2.3.1)	Self-testing (§ 2.1.2), Direct fidelity estimation (§ 2.1.4), Shadow tomography (§ 2.1.5)	Quantum state simulation, quantum channels	"Constrained" tomography (§ 2.3.2)
Verifying quantum processes	Quantum process tomography (§ 2.4), Gate-set tomography (§ 2.4)	Circuit self-testing (§ 2.1.2), Process fidelity estimation (§ 2.1.4)	Quantum circuit simulation, noise models	Randomised benchmarking (§ 2.4)

Table 2.1: A taxonomy of classical and quantum verification techniques.

2.1 Verification of quantum states

While protocols that aim to completely characterise a quantum system, such as tomography, are widely studied, they are not the only game in town when it comes to verifying quantum systems. A verification procedure is only required to test whether a system meets a certain specification, and that specification need not require a full characterisation of the system. To be more concrete, let the "model", $\mathcal{M}$, be the set of quantum states (or processes) that denotes all possible actualities of the system. The specification, then, is just a map $\mathcal{S} : \mathcal{M} \to \mathcal{R}$ to some domain $\mathcal{R}$ that labels required properties, characteristics or parameters of the system. For example, the specification might be that the system should satisfy a certain property (e.g. "is an entangled state", or "is a non-Clifford circuit"), in which case $\mathcal{R} = \{0, 1\}$, or that it should have some set of p key parameters within an acceptable range of values, in which case $\mathcal{R} \subseteq \mathbb{R}^p$. A verification protocol then is, given (possibly multiple copies of) an element $m \in \mathcal{M}$, to estimate $\mathcal{S}(m)$. As such, there are broad families of protocols that come under the umbrella of quantum system verification, but are not tomographic in nature. We will explore a representative subset of these protocols in

this section, and defer discussion of characterisation protocols until § 2.3.

2.1.1 Property testing

Quantum property testing [39, 163] refers to the instance when the domain $\mathcal{R} = \{0, 1\}$; i.e. for some subset of states $Q \subseteq \mathcal{M}$, the specification maps to 1 (and the property is "true") and for the complement $\bar{Q}$, the specification maps to 0 (the property is "false"). More generally, property testing makes use of a notion of "nearness" to having some property; i.e. we have some metric over the model $\Delta : \mathcal{M} \times \mathcal{M} \to [0, 1]$ such that:

- $m \in \mathcal{M}$ is "δ-far" from having the property if $\Delta(m, q) > \delta$, for all $q \in Q$;
- $m \in \mathcal{M}$ is "δ-close" to having the property if $\Delta(m, q) \leq \delta$ for at least one $q \in Q$.

A verifier for a particular property, given (generally multiple copies of) a particular $m \in \mathcal{M}$, must discriminate with high probability between the cases where $m \in Q$ (i.e. m has the property) or m is δ-far from Q. The quality of the protocol is typically gauged by the scaling of the number of copies needed to discriminate these cases as a function of δ.

As an example, suppose the model is just a pair of quantum states $\{\rho_0, \rho_1\}$, and the specification maps to a single bit that just labels the states: $\mathcal{S} : \rho_0 \mapsto 0$, $\mathcal{S} : \rho_1 \mapsto 1$ (i.e. the property to be tested is just the statement "the state is ρ_1"). Then the verifier for this system is just a quantum state discrimination protocol between the states ρ_0 and ρ_1 [47, 14]. This is the simplest example of the verification task that will be the basis of Chapter 3; namely, the scenario where the property to be tested is just whether a given quantum state from the model is equal to, or far from, some fixed target state. It is straightforward to show (see § 3.2.1) that if the target state is pure, then a number of copies $n = O(1/\epsilon)$ is both necessary and sufficient to test equality to the target with high probability (where we remind the reader that ϵ is the infidelity of the given state with the target state). In the case where the target state is pure, but the set of measurements available to the verifier is restricted to Pauli measurements, then the best known prior art regarding the number of required copies scales like $O(d/\epsilon^2)$, derived using either the verification protocol in [206] or in the direct fidelity estimation protocol in [84] (indeed in the latter, the verifier does not only test whether the output state is exactly equal to a known target state, but estimates the fidelity to the target with precision $\pm\epsilon$). If the target state is mixed, and the verifier has no restriction on the types of measurement that they can apply, then the scaling is less favourable than the equivalent case for pure states; for a target $\rho \in \mathcal{D}(\mathbb{C}^d)$, the number of copies scales like $O(d/\epsilon)$ [11]. Additionally, this scaling with d is unlikely to be improved; in the particular case where $\rho = \mathbb{1}/d$, we have the tight bound that $\Theta(d/\epsilon_T^2)$ copies are required [173].

There are many more properties of quantum states which may be desirable, and are testable in this framework. For example, one may test whether a given state is separable. If the model is restricted to the set of pure states, then separability is testable with $O(1/\epsilon_T^2)$ copies [106] by applying the swap test to subsystems of two copies of the given state [160]. However, if the model allows mixed states then testing for separability is much harder. As a direct consequence of a result by Childs, Harrow and Wocjan [50], it is shown in [163] that $\Omega(d^2/\epsilon_T^2)$ copies are needed to test for separability, in general. On the other hand, testing whether a state is pure is generally efficient for arbitrary states; the purity of a quantum state ρ is given by $P = \mathrm{tr}(\rho^2)$, and applying the swap test to a pair of copies of ρ accepts with probability $\frac{1}{2}(1+P)$. Therefore repeatedly applying the swap test to pairs of copies of ρ is enough to estimate the purity. If ρ is ϵ_T-far from being pure, then the swap test rejects with probability $\frac{\epsilon_T}{2}$, and so only $O(1/\epsilon_T)$ copies are needed to test for purity. One final property that we remark upon is the testing of whether a given state is a stabilizer state, or far from one (see § 3.5 for the definition of a stabilizer state). If the model consists of the set of stabilizer states over $\log d$ qubits, then the related task of determining which stabilizer state has been given to the verifier requires measuring $\Omega(\log d)$ copies, by an appeal to Holevo's theorem [116]. An algorithm by Aaronson and Gottesman [3] gave a matching upper bound of $O(\log d)$ if collective measurements on all n copies are allowed, and an upper bound of $O(\log^2 d)$ for single-copy measurements. An upper bound of $O(\log d)$ for two-copy measurements was shown by Montanaro [162] (with an algorithm based on the swap test, of a similar flavour to that used to test for separability for pure states). However, this does not preclude the notion of a tester for stabilizer states that does not scale with the number of qubits. An efficient tester for this problem was recently devised by Gross, Nezami and Walter, requiring a number of copies $O(1/\epsilon)$ [95] (independent of the number of qubits over which the stabilizer state is defined).

2.1.2 Self-testing

In most verification scenarios, we typically make some subset of the following assumptions: (i) that the measurement devices faithfully perform the POVM we expect; (ii) that we can reliably perform the same measurement across multiple trials; (iii) Markovianity (i.e. that outcomes from the experiment do not depend on previous trials); (iv) that the Hilbert space is of a fixed and known dimension; (v) that classical analysis and post-processing is perfect (vi) that locality assumptions about the device and measurements are satisfied. *Self-testing* is a field dedicated to a family of verification protocols for testing equality to a target state that do away with most of these assumptions. Specifically, self-testing protocols rely only on

assumptions (iii) and (vi) (in other words, the assumptions that freely chosen measurements are independent, and that no-signalling holds). The other assumptions can be discarded.

The most well-known self-test is of the singlet, based on the Clauser-Horne-Shimony-Holt (CHSH) inequality [61][1]. While it is widely understood that a CHSH test of the singlet (experimental imperfections aside) saturates Tsirelson's bound $S_{CHSH} = 2\sqrt{2}$ [220], the converse is also true; namely:

Theorem 1 (Singlet self-testing, Paraphrased from [183, 33])**.** *If the outcome of a CHSH test yields $S_{CHSH} = 2\sqrt{2}$, then not only was the experiment producing copies of a Bell state (up to local isometries), but the applied measurements must have been of the correct Pauli observables (up to local isometries).*

Thus by carrying out a CHSH test and analysing correlation data alone, one can verify both the generation of singlet states and the correct operation of a potentially untrusted measurement apparatus. Historically, while this result was first being considered in the context of state verification, Mayers and Yao published an alternative self-test of the singlet that relied on different observables [154] (in particular, that Alice and Bob claim to be able to measure three distinct observables, rather than two). Interestingly, the correlations observed for the Mayers-Yao test can only achieve $S_{CHSH} = \sqrt{2} + 1 < 2\sqrt{2}$ [230] and so are not able to saturate Tsirelson's bound, implying that each test is distinct. In the case where each party is able to choose from two measurement settings each with two outcomes, the full set of self-tests for the singlet is known [230].

There is a rapidly growing class of states beyond the singlet for which self-tests are known. Regarding bipartite states, it was previously known that any pure, bipartite, partially-entangled state of two qubits can be self-tested by parties with two settings and two outcomes [234], and for pairs of qutrits, both the partially-entangled states that saturate the Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality [67] can be self-tested [235], as can the maximally-entangled state [196]. However, these results have been superseded by work by Coladangelo, Goh and Scarani that demonstrate that, for any bipartite qudit state, there exists: (i) correlations that self-test that state [65]; and (ii) a generalised CHSH game that matches those correlations [64]. In the case of multi-copy verification, multiple singlets can be self-tested both sequentially and in parallel [193, 158, 167]. Regarding states with more than two parties, it is known that all graph states can be self-tested (by measuring their corresponding stabilizers) [155], as can the W state [177, 233]. Additionally, it is shown in [214] that all Dicke states,

[1] Readers unfamiliar with the CHSH inequality can find a primer in [170, §2.6].

partially-entangled GHZ states, and even all multipartite qudit states that admit a Schmidt decomposition are self-testable.

A reasonable criticism of these results, in isolation, is that it is highly infeasible to create even a singlet with perfect fidelity. However, we would hope that if we recorded a Bell violation of $S_{CHSH} = 2\sqrt{2} - \delta$, say, that the output state also has some high fidelity with the singlet. This notion of *robustness* in self-testing was initially explored for the singlet by McKague [156]. Specifically, it is shown that achieving a violation of $S_{CHSH} = 2\sqrt{2} - \delta$ lower bounds the fidelity with the singlet as

$$F(\rho, |\Phi^+\rangle) \geq 1 - \frac{1}{4}\left(9\sqrt{2}\delta - 100 \cdot 2^{\frac{1}{4}}\delta^{\frac{1}{2}} + 60 \cdot 2^{\frac{3}{8}}\delta^{\frac{3}{4}}\right). \tag{2.1}$$

While this result is a step in the right direction, it does not provide practically useful bounds on the fidelity; if $\delta = 10^{-4}$, say, then the fidelity is only lower bounded by $F(\rho, |\Phi^+\rangle) \geq 0.2$. Thankfully this bound was drastically improved, first to the bound $F \gtrsim 1 - 1.1\delta$ in [12, 235], and then to the bound $F \gtrsim 1 - 0.69\delta$ [124]. Also, robustness results exist for the W state, linear cluster states and the GHZ state [177, 124].

2.1.3 Bell tests

We have seen that, particularly in the case of two-qubit states, to verify the presence of a certain state one might perform some kind of Bell test (possibly tailored to the state of interest). We would now like to explore Bell tests in the context where: (i) the measurement devices are trusted; and (ii) we only perform the test with a finite number of copies of an output state. In a similar vein to the robustness arguments above, one may ask: in this scenario, to what accuracy can a Bell parameter be stated? And how does this accuracy in the Bell parameter translate into an accuracy in, say, the fidelity of the state with the expected state? We will see, for two-qubit states, that the outcome is broadly similar to the results we describe for tomography in § 2.3.1 - that we need at least a number of copies $O(1/\epsilon^2)$ in order to verify a two-qubit state to within fidelity $\pm\epsilon$ using a Bell test. For simplicity, we focus on a CHSH test rather than another, more complicated non-local game.

To begin with, we will first give a relationship between the required number of copies, n, and the *mean squared error* (MSE) between the observed output data from the CHSH game, and their expectations given the output state. For a set of n estimates $\tilde{X}_i$ with expectations X_i, the MSE is given by

$$\Delta^2 := \frac{1}{n}\sum_i (\tilde{X}_i - X_i)^2. \tag{2.2}$$

Then given a CHSH game carried out on n identical copies of a two-qubit state with expected parameter S_{CHSH}, we can relate n and the MSE. The following theorem is from [212], § 4.3:

Theorem 2 (Bell test MSE [212])**.** *Given a CHSH game carried out on n identical copies of a two-qubit state with expected Bell parameter S_{CHSH}, to bound the MSE by Δ^2 it is necessary for the verifier to measure a number of copies*

$$n \geq \frac{16 - |S_{CHSH}|^2}{\Delta^2}. \tag{2.3}$$

The proof of this theorem is included in both [212] and Appendix A. It remains to translate this copy complexity in terms of the MSE into a copy complexity in terms of the fidelity of the output of the experiment with a target state. One subtlety regarding the CHSH test is that Alice's and Bob's measurement settings must be "calibrated" to the target state. So if, for example, they expect copies of the state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and set up their measurement settings accordingly, then upon receipt of copies of any other Bell state (say, $|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$), they will output a Bell parameter $S_{CHSH} = 0$ despite their input being maximally entangled. Thus the verifier must have the ability to calibrate their choice of CHSH measurement settings by local rotations in order to align with the target state. A corollary to this is that, as any two-qubit state is locally equivalent to a state $|\psi_\theta\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, for some θ, we can restrict to target states of this form without loss of generality. Moreover, it is clear from Eq. 2.3 that the required number of copies is minimised by maximising $|S_{CHSH}|$, and so it is in the verifier's interest to calibrate the measurement settings to maximise the violation for a given target $|\psi_\theta\rangle$. In this instance, the MSE and the fidelity are then related by the following:

Lemma 3. *Consider a CHSH test carried out on n copies of an output state claimed to be $|\psi_\theta\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, where we must use this test in order to discriminate from a state σ such that $F(|\psi_\theta\rangle, \sigma) = 1 - \epsilon$, $0 \leq \epsilon \leq \frac{1}{2}$. Then the root MSE between these two cases is given by*

$$\Delta = |\epsilon(S_\theta - S_\phi) - \sqrt{\epsilon(1-\epsilon)}S_j| \tag{2.4}$$

for ϵ-independent parameters S_θ, S_ϕ and S_j that depend upon $|\psi_\theta\rangle$, σ, and the choice of calibration for the CHSH test. Thus there is a lower bound on the copy complexity:

$$n \geq \frac{16 - S_\theta^2}{\epsilon^2(S_\theta - S_\phi - S_j)^2} \tag{2.5}$$

copies are necessary to verify that σ is within infidelity ϵ to the target state $|\psi_\theta\rangle$ using a CHSH test.

The proof of this statement is also included in Appendix A. The salient point is that for fixed θ, the number of copies to distinguish a target state $|\psi\rangle$ from a state ϵ away using a CHSH test scales like $O(1/\epsilon^2)$, in analogy with the bounds for quantum state tomography in § 2.3.1.

2.1.4 Direct fidelity estimation

In contrast to the above, an example of a verification protocol that does not fit neatly into the property testing framework is the *direct fidelity estimation* protocol in [84]. The name conveys most of the premise of this verification scenario: given multiple copies of some state ρ, and the ability to apply Pauli measurements on individual qubits, the verifier must estimate the fidelity with some pure target state $|\psi\rangle$ as accurately as possible.

Let $\rho \in \mathcal{D}(\mathbb{C}^{2^N})$; we denote a particular N-fold tensor product of Pauli observables as σ_i, for $i = 1 \ldots 4^N$. The goal of the verifier is to produce an estimate $\tilde{F}$ of the true fidelity $F = \langle\psi|\rho|\psi\rangle$, up to accuracy $\pm\epsilon$. The protocol is randomised, and so we can only claim to satisfy this accuracy condition with probability $1-\delta$. Given these requirements, the direct fidelity estimation protocol is the following, where parameters will be chosen depending on $|\psi\rangle$:

Protocol Direct fidelity estimation

1: **for** $i = 1$ to ℓ **do**
2: Randomly pick Pauli measurement setting σ_i with probability p_i
3: **for** $j = 1$ to m_i **do**
4: Measure σ_i for a copy of ρ and store outcome as a_{ij}
5: Produce estimate $X_i = \frac{C_i}{m_i} \sum_{j=1}^{m_i} a_{ij}$
6: Output fidelity estimate $\tilde{X} = \frac{1}{\ell} \sum_{i=1}^{\ell} X_i$.

Theorem 4 (Direct fidelity estimation [84])**.** *Suppose the verifier makes the following choices for the parameters ℓ, m_i, C_i and p_i:*

$$\ell = \frac{8}{\epsilon^2 \delta}; \quad p_i = \frac{|\langle\psi|\sigma_i|\psi\rangle|^2}{d}; \quad m_i = \frac{d\delta}{4|\langle\psi|\sigma_i|\psi\rangle|^2} \log\frac{4}{\delta}; \quad C_i = \frac{1}{\langle\psi|\sigma_i|\psi\rangle}. \tag{2.6}$$

Then it is guaranteed that

$$Pr[|\tilde{X} - F| \geq \epsilon] \leq \delta. \tag{2.7}$$

The proof of this theorem is in [84] and is included in Appendix A. It is worth remarking that, unlike tomography, the necessary number of *measurement settings* is independent of d; it is just $\ell = \frac{8}{\epsilon^2\delta}$. However, the necessary number of *measurements* does depend on d; it is bounded by

$$\sum_i \mathbb{E}(m_i) \leq 1 + \frac{8}{\epsilon^2\delta} + \frac{8d}{\epsilon^2} \log\frac{4}{\delta}. \tag{2.8}$$

While this dependence is exponential in the number of qubits, it is still polynomially better than the scaling that is inherent to quantum state tomography. It is also worth mentioning that, for many states of interest in quantum information, the scaling of this protocol may be markedly better than this worst-case bound:

Lemma 5. *Suppose it is guaranteed that, for a fixed target state $|\psi\rangle$ and for any choice of Pauli σ_i, either: (a) $\langle\psi|\sigma_i|\psi\rangle = 0$, or (b) $|\langle\psi|\sigma_i|\psi\rangle| \geq \alpha$, for some threshold α. Then*

$$\sum_i \mathbb{E}(m_i) \leq O\left(\frac{1}{\alpha^2\epsilon^2}\log\frac{1}{\delta}\right). \tag{2.9}$$

In particular, for stabilizer states we have that $\alpha = 1$, and so the copy complexity does not grow with d. For W states, it can be shown that $\alpha = \frac{1}{\log d}$, and so the scaling is quadratic in the number of qubits.

2.1.5 Shadow tomography

A verification task that sits close to the boundary between verification and characterisation is *shadow tomography* [2]. In this instance, the specification is not to output a single bit, as in property testing, or a single real number, as in fidelity estimation, but to output a list of real numbers, $\{b_1, b_2 \ldots b_m\}$. Given (multiple copies of) a state $\rho \in \mathcal{D}(\mathbb{C}^d)$ and a list of m observables $E_1 \ldots E_m$, the parameter b_i should give a close approximation to the expectation value $\mathrm{tr}(E_i\rho)$, for all i, using as few copies of ρ as possible. This task is weaker than full tomography; the set $\{E_i\}$ need not be informationally complete (i.e. it may not be possible to output a complete classical description of a state that approximates ρ using the set of b_i's alone). A corollary is that, if the verifier were to carry out full tomography on ρ, they could trivially output a set $\{b_1 \ldots b_m\}$; using the bounds for sample-optimal tomography in § 2.3.1, it is sufficient in this instance to measure $O\left(d^2\right)$ copies. On the other hand, if the verifier were to just apply the measurement E_i to copies of ρ, they could output the full list of approximations using $O(m)$ copies. The result in [2] notably derives an upper bound in copy complexity that is a significant improvement over both of these naïve approaches.

Theorem 6 (Shadow tomography [2])**.** *Given n copies of some d-dimensional mixed state, $\rho^{\otimes n}$, and a description of two-outcome observables E_i, $i = 1 \ldots m$, there is an explicit protocol that outputs a list $\{b_1 \ldots b_m\}$ such that, with probability $1-\delta$,*

$$|b_i - \mathrm{tr}(E_i\rho)| \leq \epsilon, \ \forall i. \tag{2.10}$$

Moreover, we have a copy complexity

$$n = \tilde{O}\left(\frac{1}{\epsilon^5}\log m \log d \log\frac{1}{\delta}\right), \tag{2.11}$$

where the $\tilde{O}$ notation hides logarithmic factors of ϵ, and doubly logarithmic factors of d and m.

The spirit of the approach is to apply a form of "gradient descent"; we have some hypothesis state ρ_t, initially at $\rho_0 \propto \mathbb{1}$, and we seek to update by descending along the direction that represents the worst-case for the verifier; i.e. we look for and measure observables E_i where $|\mathrm{tr}(E_i\rho_t) - \mathrm{tr}(E_i\rho)|$ is big. Finding informative E_i's for the verifier relies on the *gentle search procedure*, based on a protocol by Harrow, Lin and Montanaro [105]; this protocol decides whether an observable from the set accepts with high probability, or that all observables accept with low probability, given the promise that one or the other is true.

2.1.6 Verification and hypothesis testing

Ultimately, the procedure for verifying the existence of a particular quantum state $|\psi\rangle$ is nothing more than a binary hypothesis test; we perform measurements on multiple copies of the output state, collect data and construct statistics, then on the basis of these statistics we either accept a null hypothesis H_0 that the output state was indeed $|\psi\rangle$, or we reject in favour of an alternative hypothesis H_1, that the output state is far from $|\psi\rangle$.

In a binary hypothesis test between hypotheses H_0 and H_1, there are two ways in which we can err, commonly called "Type I" and "Type II" errors. These are, respectively,

$$\text{Type I}: \quad \Pr[\text{Guess } H_1 | H_0] := \alpha, \tag{2.12}$$

$$\text{Type II}: \quad \Pr[\text{Guess } H_0 | H_1] := \beta. \tag{2.13}$$

In general, in designing an effective hypothesis test there will be a trade-off between the relative magnitude of these types of error; they cannot be arbitrarily decreased simultaneously. It is typically down to the experiment designer to determine in which ratio one can tolerate Type I versus Type II errors. In *symmetric* hypothesis tests, the Type I and Type II errors are treated equally, and the designer seeks to minimise them both at the same rate. However, in many scenarios these errors are valued differently. In the context of verifying quantum states, for example, misattributing the state as $|\psi\rangle$ when the output is actually far from $|\psi\rangle$ is likely to result in the failure of the entire quantum computation, whereas asserting that the state is far from $|\psi\rangle$ when the state is actually $|\psi\rangle$ just leads to recalibration or further verification. It can be reasonably argued, therefore, that we should seek to minimise the former as much as possible, even at the potential expense of marginally increasing the latter. This is the remit of *asymmetric* hypothesis tests, wherein the goal is to maximise the rate at which one of these errors decreases (in this case, the Type II error) for increasing trials, given a fixed, tolerable upper bound on the other (in this case, the Type I error).

The study of both symmetric and asymmetric hypothesis tests in quantum information is well-established (see [9] for a comprehensive review). In the context of symmetric tests we wish to minimise a (potentially weighted) sum of the Type I and Type II errors; asymptotically, this sum decays exponentially with the number of trials. In the case of discriminating two discrete, classical probability distributions, the optimal exponent is given by the *Chernoff information* [49]:

Theorem 7 (Chernoff information [49])**.** *Let $X_1 \dots X_n$ be drawn iid from either p (hypothesis H_0) with probability π_p, or q (hypothesis H_1) with probability π_q; then the probability of error is $P_{err} = \alpha\pi_p + \beta\pi_q$. This decreases exponentially with the number of trials: $P_{err} \sim \exp\{-nC(p,q)\}$, and the optimal error exponent is given by*

$$\xi_C(p,q) := \max_s C(p,q) = -\log\left[\inf_{0\leq s\leq 1} \sum_i p(i)^{1-s} q(i)^s\right]. \tag{2.14}$$

In the quantum case of discriminating n copies of two finite-dimensional density matrices $\rho^{\otimes n}$ and $\sigma^{\otimes n}$, there is a remarkably similar equation that governs the optimal exponent, naturally called the *quantum Chernoff information* [8, 171]:

Theorem 8 (Quantum Chernoff information [8, 171])**.** *Let $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ be two density matrices of finite but unknown dimension, given to a verifier with prior probabilities π_ρ and π_σ. Then the Helstrom bound for the probability of error is*

$$P_{err} = \frac{1}{2}\left(1 - \|\pi_\rho \rho^{\otimes n} - \pi_\sigma \sigma^{\otimes n}\|_1\right). \tag{2.15}$$

This probability of error decreases exponentially with the number of trials, $P_{err} \sim \exp\{-nQ(\rho,\sigma)\}$, and the optimal error exponent is given by

$$\xi_Q(\rho,\sigma) := \max_s Q(\rho,\sigma) = -\log\left[\inf_{0\leq s\leq 1} \operatorname{tr}(\rho^{1-s}\sigma^s)\right]. \tag{2.16}$$

In the asymmetric case, the statement of optimality of classical hypothesis tests is given by the *Chernoff-Stein lemma*, which is the jumping-off point for the results in Chapter 3 and will be our focus in § 3.2.2 (we refer the reader there for a full statement of the lemma). This lemma states that the probability of error also decays exponentially, with optimal exponent in the classical case given by the *relative entropy* (also known as the *Kullback-Leibler divergence*), denoted $D(\cdot\|\cdot)$. For the case of discrete probability distributions p and q, it is defined as

$$D(p\|q) = \sum_i p(i) \log \frac{p(i)}{q(i)}. \tag{2.17}$$

There is a quantum generalisation of the relative entropy, naturally called the *quantum relative entropy* [113]:

$$D_Q(\rho\|\sigma) = \operatorname{tr}[\rho(\log\rho - \log\sigma)]. \tag{2.18}$$

The Chernoff-Stein lemma has also been extended to the quantum case; first by Hiai and Petz [113], who showed that the quantum relative entropy upper bounds the optimal exponent for an asymmetric hypothesis test between two states, and then by Ogawa and Nagaoka [176], who showed that this bound is tight.

2.2 Verification of quantum processes

We now briefly highlight analogous verification procedures for quantum processes as those discussed for quantum states above.

Property testing of quantum processes is a little more subtle than for states, as we have more freedom to define plausible verifiers. For example, in testing a unitary U, we must consider the available input states and measurement settings, as well as whether the verifier has access to derivative gates such as U^{-1} or c-U (the "controlled" form of U). On the other hand some protocols for property testing processes follow neatly from, and inherit the copy complexity of, their counterparts for property testing quantum states due to the Choi-Jamiołkowski isomorphism [53, 120]. In particular, testing whether a unitary is a particular target unitary (up to a global phase) follows directly from the equivalent test for states, and requires $O(1/\epsilon)$ uses of the unitary. The test for locality of the unitary (i.e. that it is of the form $U = \bigotimes_i U_i$) is derived directly from the swap test for separability of states [160]. Membership of the Clifford group follows from the test for inclusion in the set of stabilizer states in [95]. The most efficient (and provably optimal) algorithm for learning a particular Clifford operation, given access to the operation and its inverse, has a copy complexity of $O(1/\epsilon)$, independent of the size of the Hilbert space over which the operation acts [224].

Results are also known about the ability to self-test quantum gates and circuits, rather than quantum states. The earliest result in this setting is due to van Dam et al. [71].[2] The premise is analogous in spirit to deductive verification: given a quantum channel Λ, we must construct a set of necessary conditions based on outcome probabilities that verify whether Λ is the target unitary. van Dam et al. derive a set of necessary conditions for a universal gate set. Moreover, this scheme is shown to be robust; that is, if the conditions are close to being satisfied then the channel is close to Λ. A protocol that is closer in spirit to true self-testing was later developed by Magniez, Mayers, Mosca and Ollivier [148]. The protocol only makes

[2]While this work historically comes under the umbrella of self-testing, it assumes a verifier with more control than just access to classical data corresponding to untrusted measurement outcomes; the verifier must also have the ability to both prepare probe states in the computational basis, measure in the computational basis, assume the system is of fixed and known dimension, and apply the given operator as many times as they like. As such, it has more in common with the fidelity estimation protocol in § 2.1.4 than the self-testing protocols in § 2.1.2.

the assumptions that are necessary for self-testing, plus the additional assumption that the circuits only operate on states with real-valued amplitudes (for mathematical expediency). The crux of the proposal is to generate Bell pairs and to distribute each qubit to one of two copies of the circuit; then, by running experiments with different parts of the circuit removed and performing a self-test on the states at the output, to restrict the actions of an adversarial device. They are also able to demonstrate robustness in this setting.

There is also an analogous notion of direct fidelity estimation for quantum processes [115, 192, 153]. In these scenarios, we assume that the verifier has access to some black-box unitary that is purported to be U, but actually operates with the channel Λ. The verifier is able to generate copies of some input states to Λ, and then measure at the output. The goal is to estimate some fidelity between Λ and U. Research in this direction has aimed at minimising the number of different probe states that the verifier must prepare to produce tight bounds on the process fidelity. The minimum ensemble size was recently shown to be of size d [153]. However, to date there is not a clear analysis that explores optimality in the sense of copy complexity, and the nature of optimal protocols in this setting.

2.3 Characterisation of quantum states

Characterisation of quantum systems is probably the most well-studied verification task in quantum information. In this instance, we have a specification that is a full mathematical description of the system in question (be it a quantum state or process). The goal of the verifier is to probe the system, to provide an estimate of every parameter describing the system with respect to the model, and to estimate whether it is close to the specification, in some sense.

We will focus on the characterisation of quantum states in this section; we defer discussion of quantum processes until § 2.4.

2.3.1 Quantum state tomography

In the context of quantum states, characterisation tasks are known as *quantum state tomography*. Characterisation of quantum states is typically a time-consuming and computationally difficult process. For example, tomographic reconstruction of a state of 8 ions required taking $\sim 650{,}000$ measurements over 10 hours, and a statistical analysis that took far longer [102]; verification of a few-qubit photonic state is similarly challenging [42, 132]. This is also the case in tomography of continuous-variable systems [145, 15, 7]. One may instead resort to non-tomographic methods to characterise the output state of an experiment, but such

methods typically either: (a) assume that the output state is within some special family of states, for example in compressed sensing [85, 96] or matrix product state tomography [69]; or (b) extract only partial information about the state, such as when estimating entanglement witnesses [217, 218]. We will return to these "constrained" tomography protocols in § 2.3.2. Unlike some of the verification protocols discussed above, in this setting it is common to make quite strong assumptions about the ability of the verifier. In particular, we make all of the same assumptions listed at the start of § 2.1.2.

Most tomographic protocols assume single-copy measurements only, where each copy is consumed or destroyed when the measurement is completed (this is in contrast with the "sample-optimal" tomographic schemes that we discuss in § 2.3.3). For a single copy of ρ, we make a choice of POVM indexed by k, Π_k. We let this POVM have b_k potential outcomes, indexed by $b = 1 \dots b_k$; the POVM elements of Π_k are labelled Π_k^b such that $\sum_{b=1}^{b_k} \Pi_k^b = \mathbb{1}$. Then the Born rule for measurement outcomes is $\Pr(b|\Pi_k, \rho) = \mathrm{tr}(\Pi_k^b \rho)$. If we have n total copies of ρ, denote n_k as the number of copies over which Π_k is measured. Let c_{kb} be the number of times that outcome b is recorded when measuring Π_k on n_k copies; the relative frequency of this outcome for this POVM is then denoted $f_{kb} \coloneqq c_{kb}/n_k$. It is clear that $\mathbb{E}(f_{kb}) = \mathrm{tr}(\Pi_k^b \rho)$, but it is unlikely that the actual value of f_{kb} satisfies this equality for a finite number of samples.

Suppose we wish to characterise the state up to trace distance $O(\epsilon_T)$; then we note an argument first pointed out in [85]: characterising up to $O(\epsilon_T/\sqrt{d})$ in Frobenius norm guarantees that the trace distance is at most $O(\epsilon_T)$. To characterise an arbitrary state $\rho \in \mathcal{D}(\mathbb{C}^d)$, one has to write down $d^2 - 1$ real parameters and so $\Omega(d^2)$ measurement settings are necessary. Then, to attain accuracy $O(\epsilon_T/\sqrt{d})$ in Frobenius norm by linear inversion tomography (as seen below), it suffices to estimate the expectation value for each setting up to additive error $O(\epsilon_T/d)$, which requires $O(d^2/\epsilon_T^2)$ copies per setting, by Hoeffding's inequality [114]. Therefore for the non-adaptive estimators we introduce below, $n = O(d^4/\epsilon_T^2)$ copies are needed to guarantee that the estimate is within trace distance $O(\epsilon_T)$ of the true state.

The simplest tomographic estimate that one could consider is known as *linear inversion*. Suppose ρ has some decomposition

$$\rho = \frac{\mathbb{1}}{d} + \frac{1}{2}\mathbf{r} \cdot \sigma, \tag{2.19}$$

where the vector σ is a list of $d^2 - 1$ Hermitian, traceless and orthogonal matrices[3] and the elements of $\mathbf{r}$ are a set of real parameters specified by $r_i = \mathrm{tr}(\rho\sigma_i)$. Then

[3]Here we have used the notation σ to indicate Pauli matrices, but any informationally-complete set of Hermitian and traceless matrices will do.

POVM elements can also be decomposed in this fashion:

$$\Pi_k^b = s_{kb}^0 \mathbb{1} + \mathbf{s}_{kb} \cdot \sigma, \tag{2.20}$$

where $s_{kb}^0 = \mathrm{tr}(\Pi_k^b)/d$ and $s_{kb}^i = \mathrm{tr}(\Pi_k^b \sigma_i)/2$. Then outcome probabilities can be rewritten

$$\Pr(b|\Pi_k, \rho) = \mathrm{tr}(\Pi_k^b \rho) = s_{bk}^0 + \mathbf{s}_{bk} \cdot \mathbf{r}. \tag{2.21}$$

A linear inversion estimator simply equates this expression with the observed relative frequencies:

$$s_{bk}^0 + \mathbf{s}_{bk} \cdot \mathbf{r} = f_{bk}, \ \forall b, k. \tag{2.22}$$

By vectorising, solving this is equivalent to solving the corresponding matrix equation for $\mathbf{r}$:

$$S\mathbf{r} = \mathbf{f}. \tag{2.23}$$

There is no reason to expect the dimension of $\mathbf{r}$ to be the same as the dimension of $\mathbf{f}$, and so S is not square; in this instance one can use the Moore-Penrose pseudoinverse to output

$$\mathbf{r} = (S^\dagger S)^{-1} S^\dagger \mathbf{f}. \tag{2.24}$$

A linear inversion estimate ρ_L is constructed by evaluating this expression and plugging the resultant $\mathbf{r}$ into Eq. 2.19. This method has some useful operational properties. Firstly, it is comparatively transparent what must be calculated in order for ρ_L to be produced. Secondly, for quantities derived from ρ_L, such as the fidelity with a target or an entanglement witness, it tends to produce estimates that are significantly less biased than common alternatives [201]. However, it has a deficiency: there is no guarantee that ρ_L corresponds to a physical density matrix. In practice, linear inversion estimates commonly report negative eigenvalues [72].

The most common and widely-adopted alternative is *maximum likelihood tomography* [117]. The likelihood function $\mathcal{L}(\rho)$ evaluates the probability of observing the full data set $c = \{c_{kb}\}$ given a particular ρ:

$$\mathcal{L}(\rho) := \Pr(c|\rho) = \prod_{k,b} \mathrm{tr}(\Pi_{kb}\rho)^{c_{kb}} = \prod_{k,b} \mathrm{tr}(\Pi_{kb}\rho)^{f_{kb}/n}. \tag{2.25}$$

It is more practical to consider the *log-likelihood*:

$$\log \mathcal{L}(\rho) = \frac{1}{n} \sum f_{kb} \log \mathrm{tr}(\Pi_{kb}\rho). \tag{2.26}$$

The maximum-likelihood estimate ρ_{MLE} is then given by

$$\rho_{MLE} = \arg\max_\rho \log \mathcal{L}(\rho) = \arg\max_\rho \frac{1}{n} \sum f_{kb} \log \mathrm{tr}(\Pi_{kb}\rho),$$
$$\text{subject to } \rho \geq 0, \ \mathrm{tr}\,\rho = 1. \tag{2.27}$$

The advantage of this estimate over ρ_L is that the constraints guarantee that a physical density matrix is produced (i.e. one with non-negative eigenvalues). On the other hand, calculating ρ_{MLE} is incredibly expensive computationally [102, 202]. Additionally, ρ_{MLE} is very likely to be rank-deficient; that is, to have some set of eigenvalues that vanish entirely [25]. While this is physically allowable, it is statistically problematic; the only error bars consistent with this point estimate are ones that vanish entirely. Hence if $|u\rangle$ is an eigenstate of ρ_{MLE} with zero eigenvalue, then we must conclude that measuring $\mathrm{tr}(|u\rangle\langle u|\rho)$ will yield outcome zero, with complete confidence; however, this is preposterous given that we have taken a finite amount of data to construct ρ_{MLE}. The issue is that, in some sense, maximum likelihood estimation takes unphysical estimates, i.e. estimates outside the space of valid density matrices, and projects them onto the closest physical state - which invariably is a state on the boundary [25]. These boundary states are typically rank-deficient. This could also be argued to be the source of bias in MLE, over linear inversion: artificially increasing negative eigenvalues so that the state becomes positive means that large eigenvalues must be decreased if the trace is to remain fixed, and these eigenvalues contribute the most to quantities like the fidelity [25].

To attempt to remedy these issues, one may resort to Bayesian techniques[4] [25, 118]. Using Bayes' Rule, we have a posterior distribution on ρ given data c as

$$\Pr(\rho|c)d\rho = \frac{\Pr(c|\rho)\Pr(\rho)d\rho}{\Pr(c)} = \frac{\mathcal{L}(\rho)\Pr(\rho)d\rho}{\Pr(c)}, \tag{2.28}$$

where the prior distribution $\pi(\rho) = \Pr(\rho)d\rho$. The *Bayesian mean estimate* is given by

$$\rho_{BME}(c) = \mathbb{E}_\rho[\rho|c] = \int d\rho\rho\Pr(\rho|c). \tag{2.29}$$

This estimate has some advantageous features over its frequentist counterparts: (i) it does not lead to rank-deficient estimates; (ii) it yields well-defined and meaningful confidence regions around the point estimate; (iii) the estimator is optimal with respect to a family of metrics called *operational divergences* [25, 93]. On the other hand, it is: (i) even more computationally taxing to compute than ρ_{MLE}; (ii) it is not immediately obvious how to pick an appropriate prior (that is, one that is at least somewhat informative, without overstating the confidence in the estimate before the experiment begins). In general, there is no clear choice of estimator in quantum state tomography; none are without flaws. A verifier must make an informed and transparent choice based on the context of the experiment.

In contrast to the estimators above, it may also be beneficial to the verifier to consider *adaptive* tomographic protocols; that is, protocols that allow a verifier to

[4] It is worth noting that MLE can be seen as a particular instance of a Bayesian estimator. If we specify a uniform prior over all states ρ, then the MLE is equivalent to the *maximum a posteriori* estimate - the modal point of the posterior distribution.

modify the choice of future measurement settings based on past measurement data. The advantage here is that, in principle, it takes far fewer copies to produce an estimate to within an infidelity ϵ ($O(1/\epsilon)$ rather than $O(1/\epsilon^2)$ for conventional tomography) [149, 209, 210]. However, adaptive protocols are difficult to treat analytically and the evidence for this advantage is purely numerical. While this advantage in copy complexity may look promising, it may be the case that the computational cost of computing the next measurement setting is particularly taxing, given the data. It is not implausible that this added time penalty incurred by adaptively computing measurement settings negates any advantage in copy complexity, in real life experiments.

2.3.2 "Constrained" tomography

The most common approach to negate the impact of the exponential scaling of the copy complexity with the number of qubits is to constrain the state ρ to lie in a subset $M \subset \mathcal{D}(\mathbb{C}^d)$. It is debatable where to place such methods in the taxonomy of verification methods in Table 2.1. On the one hand, if we have a strict guarantee that the tested state cannot stray outside the subset M, then "constrained" tomography of this type can be thought of as *model checking* in the same sense as conventional tomography, but where the model only contains states in M. On the other hand, if there is any non-zero probability for ρ to stray outside M then the model may contain all states in $\mathcal{D}(\mathbb{C}^d)$ and we are in the realm of *testing*, where we have no strict guarantee that the protocol that we implement is sufficient to faithfully characterise the state. We choose the latter, as (a) guarantees of this type are implausible in practice and (b) most protocols we consider below explicitly treat cases where ρ is close to, but not strictly an element of, M [96, 85]. We will now highlight a pair of notable protocols of this type.

We have already mentioned a scenario of this type; when M is the set of stabilizer states of $\log d$ qubits. As noted in § 2.1.1, it is intuitive from Holevo's theorem that we expect, given a stabilizer state of $\log d$ qubits, to consume at least $\Omega(\log d)$ copies to identify it. Protocols by Aaronson and Gottesman [3] and Montanaro [162] show that this bound is tight; that $O(\log d)$ copies are sufficient. The protocol consists of taking two copies of the unknown stabilizer state, and performing Bell basis measurements on pairs of corresponding qubits.

One can also make some headway in alleviating the ailments of tomography by assuming that ρ is low rank; i.e. that it has at most $k \ll d$ non-zero eigenvalues. Protocols under this assumption are commonly called *quantum compressed sensing* [96, 85]. The nature of the protocols in [96, 85] are very similar to the direct fidelity estimation protocol in § 2.1.4. Given a copy of ρ, the verifier randomly picks

one of the d^2 Pauli settings and measures. To reconstruct the state to within trace distance ϵ_T, it is shown in [96, 85] that it is sufficient for the verifier to randomly choose $O(kd\log d)$ settings from the d^2 that are possible, and to measure each setting on $O(kd)$ copies. Thus the protocol has a copy complexity of $O(k^2d^2\log d)$. Moreover, for protocols that use single-copy Pauli measurements, this bound is almost tight; it is necessary to measure $\Omega(k^2d^2/\log d)$ copies to guarantee a reconstruction up to constant trace distance. While the most fitting scenario for compressed sensing protocols are when ρ is guaranteed to be of low rank, it may also be useful when the available verification time is very short. Short verification times might mean that a conventional tomographic protocol does not have time to measure every setting from an informationally-complete set, or to only measure them on a few copies such that the expectation estimates are very poor. In this instance, it might be more efficient to precisely characterise a smaller subset of settings, and to output a low rank estimate by compressed sensing. [85] gives compelling numerical evidence that compressed sensing estimates may handily outperform maximum-likelihood estimation in this setting.

Schemes with reduced copy complexity also exist for other classes of states, such as matrix product states [69] and permutationally invariant states [219].

2.3.3 Sample-optimal tomography

One may ask whether the protocols in § 2.3.1 and § 2.3.2 represent the ultimate efficiency in copy complexity with respect to state characterisation, even with verifiers capable of carrying out more general measurements than, for example, Pauli measurements. The copy complexity of "sample-optimal" tomography schemes has been studied by Haah, Harrow, Ji, Wu and Yu [100], and by O'Donnell and Wright [173, 174, 175]. In this setting one may consider protocols that allow *collective* measurements, that is, any valid measurement that acts on $\rho^{\otimes n}$; or *single-copy* measurements, where the verifier is only given access to a single ρ at a time (that is consumed on measurement). A summary of both necessary and sufficient copy complexities for these scenarios are shown in Table 2.2.

It is noteworthy that most of the bounds in Table 2.2 are quite tight; in general, the poor scaling with d in all of the above tomographic settings are both explicitly calculable and completely unavoidable. Conversely, while a verifier with the ability to apply arbitrary collective measurements seems quite powerful, it only gives a quadratic advantage in d over a verifier that is constrained to only measure Paulis on single copies (for arbitrary states).

Meas. type	≥/≤	Pure states	Rank-k states	Arb. states
Collective	≥	$\Omega\left(\frac{d}{\epsilon}\right)$ [108]	$\Omega\left(\frac{kd}{\epsilon_T^2 \log \frac{d}{k\epsilon_T}}\right)$ [100]	$\Omega\left(\frac{d^2}{\epsilon_T^2}\right)$ [100]
	≤	$O\left(\frac{d}{\epsilon}\right)$ [108]	$O\left(\frac{kd}{\epsilon} \log \frac{d}{\epsilon}\right)$ [100] $O\left(\frac{kd}{\epsilon_T^2}\right)$ [173] $O\left(\frac{kd}{\epsilon}\right)$ [175]	$O\left(\frac{d^2}{\epsilon} \log \frac{d}{\epsilon}\right)$ [100] $O\left(\frac{d^2}{\epsilon_T^2}\right)$ [173] $O\left(\frac{d^3}{\epsilon}\right)$ [175]
Single-copy	≥	$\Omega\left(\frac{d}{\epsilon^2} \log \frac{1}{\epsilon}\right)$ [100]	$\Omega\left(\frac{k^2 d}{\epsilon^2} \log \frac{1}{\epsilon}\right)$ [100]	$\Omega\left(\frac{d^3}{\epsilon_T^2}\right)$ [100]
	≤	$O\left(\frac{d}{\epsilon_T^2}\right)$ [131, 96, 85]	$O\left(\frac{k^2 d}{\epsilon_T^2}\right)$ [131, 96, 85]	$O\left(\frac{d^3}{\epsilon_T^2}\right)$ [131, 96, 85]
Pauli	≥	$\Omega\left(\frac{d^2}{\epsilon_T^2 \log d}\right)$ [85]	$\Omega\left(\frac{k^2 d^2}{\epsilon_T^2 \log d}\right)$ [85]	$\Omega\left(\frac{d^4}{\epsilon_T^2 \log d}\right)$ [85]
	≤	$O\left(\frac{d^2}{\epsilon_T^2} \log d\right)$ [96, 85]	$O\left(\frac{k^2 d^2}{\epsilon_T^2} \log d\right)$ [96, 85]	$O\left(\frac{d^4}{\epsilon_T^2}\right)$ [96, 85]

Table 2.2: A breakdown of known upper and lower bounds in copy complexity for sample-optimal tomography. "≥" and "≤" respectively denote necessary (lower) and sufficient (upper) bounds in copy complexity.[5]

2.3.4 Confidence estimates in quantum state tomography

While constructing a point estimate of a state ρ alone given some measurement data is a well-motivated request, there is no guarantee that the point estimate is exactly correct given a finite data set, nor any a priori notion of how "good" this estimate is. For the point estimate to be meaningful, some generalised notion of "error bars" are required to quantify the accuracy of the estimate. We now outline established approaches to this problem in quantum state tomography.

The results we have discussed on compressed sensing [96, 85], direct fidelity estimation [84] and sample-optimal tomography [100, 173, 175] give a notion of an error bar in the form of a certificate: after measuring a fixed number of copies of the state, there is a guarantee that the point estimate is within some fixed distance of the true output state (although only the direct fidelity estimation protocol gives this certificate precisely, including constant factors). Such certificates are relatively simple to state and to computationally produce, and protocols of this type will be the focus in Chapters 3 and 4. On the other hand, error bars defined in this way have some immediate drawbacks [26]: (i) the "error ball" drawn around the estimate may contain nonphysical states (which is increasingly likely if the point estimate is close to the boundary of allowable states, see § 4.3); (ii) the ball must be centred on the estimate; (iii) the ball is of a fixed (ellipsoidal) shape. Error balls drawn this way could be both unrepresentative of the data and suboptimal in terms of size and

[5]The notation "O" and "Ω" can be thought of as asymptotic, approximate versions of $\leq$ and $\geq$, respectively. In particular, for two functions $f(n)$ and $g(n)$, $f(n) = O(g(n))$ if there is a positive real number a and integer n_0 such that, for all $n \geq n_0$, $f(n) \leq a g(n)$. An analogous definition holds for Ω, where instead $f(n) \geq a g(n)$.

shape of confidence region. For non-Bayesian protocols such as maximum likelihood estimators, there are alternatives; for example, the *confidence regions* defined by Blume-Kohout [26] and Christandl and Renner [54]. While such estimates are well-defined and statistically rigorous, they also have drawbacks: (i) they don't produce a reportable point estimate; (ii) there is no analytically tractable recipe to extract a confidence region at a given confidence, for an arbitrary data set; (iii) numerical methods to extract the confidence region are computationally expensive. However, simpler numerical approximations have also been studied [79].

In the Bayesian setting, this notion of "generalised error bars" is conceptually clearer; the verifier has some posterior distribution given the measurement data, and can use this to construct a Bayesian *credible region* that matches the frequentist notion of a confidence interval [81]. In particular, let the measured data be c and the state ρ. For a region R, let $\mathbb{1}_R(\rho)$ be the indicator function for the region R; that is, $\mathbb{1}_R(\rho) = 1$ if $\rho \in R$ and is zero otherwise. Then a region R is *α-credible* if

$$\Pr[\rho \in R|c] = \mathbb{E}[\mathbb{1}_R(\rho)] \geq 1 - \alpha. \tag{2.30}$$

While the notion is simple to state, finding credible regions has the same drawbacks as generating point estimates in Bayesian tomography: it is highly dependent on the choice of prior, and is computationally difficult to produce. However, progress has been made in both of these directions; both in the sense of developing insightful priors [91] and in developing techniques [92] and off-the-shelf software [93] that make Bayesian tomography more tractable.

Finally, we would like to highlight one particular approach to generating confidence intervals in quantum state tomography, as it will feature in our discussions on state verification in Chapter 3 and fidelity estimation in Chapter 4. This approach is called *precision-guaranteed tomography* [213]. It is a protocol that derives error bars in the "certificate" sense described above, by certifying that the output state and the target state must lie within some distance defined by the measurement data. The point estimate in this protocol is derived slightly differently to the MLE above, in that physicality of the density matrix is enforced in two stages; first, that the estimate must be Hermitian and have unit trace, and then that it is positive. Such an estimate is called an *extended norm minimisation* estimate, or ρ_{ENM}. Specifically,

$$\rho_{ENM} = \underset{\substack{\rho \\ \rho \geq 0}}{\arg\min} \|\rho - \rho'\|_2; \tag{2.31}$$

where

$$\rho' = \underset{\substack{\sigma \\ \mathrm{tr}(\sigma)=1, \sigma^\dagger = \sigma}}{\arg\min} \|\mathbf{r}(\sigma) - \mathbf{f}\|_2, \tag{2.32}$$

using the notation from Eq. 2.23. Given this estimator, the following holds:

Theorem 9 (Precision-guaranteed tomography [213])**.** *Consider a tomographic protocol that takes n copies of a target state $\rho \in \mathcal{D}(\mathbb{C}^d)$ and applies an informationally complete set of measurement settings $\{\Pi_k^b\}$. Let $\Delta(\cdot,\cdot)$ be a loss function between states (such as the infidelity). Then*

$$Pr\left[\Delta(\rho_{ENM},\rho) > \epsilon\right] \leq 2\sum_{\alpha=1}^{d^2-1} \exp\left(-\frac{b}{c_\alpha}\epsilon^2 n\right), \tag{2.33}$$

where

$$b = \begin{cases} \frac{8}{d^2-1} & \text{if } \Delta \text{ is the Hilbert-Schmidt distance} \\ \frac{16}{d(d^2-1)} & \text{if } \Delta \text{ is the trace distance} \\ \frac{4}{d(d^2-1)} & \text{if } \Delta \text{ is the infidelity} \end{cases} \tag{2.34}$$

and c_α is a constant dependent on the choice of measurement settings.

This result is enough, for example, to give a certification of the fidelity of the output state from an experiment to a target state, given a state tomography protocol constructed from Pauli measurements in the conventional way (including constant factors). We will make use of this result in Chapter 3.

2.4 Characterisation of quantum processes

The earliest example of a protocol for process characterisation is *quantum process tomography* (QPT) [55, 185]. QPT and state tomography are intimately related. Given a verifier capable of preparing perfect input states to a process Λ and to collect measurement data by measuring with perfect precision, analysis by QPT yields an analogous expression to that for state tomography in Eq. 2.21, and the verifier solves for Λ in an identical way (e.g., by linear inversion or MLE). Process tomography, then, inherits all of the same assumptions and drawbacks as state tomography. Two are particularly acute: (i) the channel is described by $O(d^4)$ real parameters, and so an analogous argument to that for quantum states would imply that, if the verifier is restricted to single-copy, two-outcome measurements, then $O(d^4)$ total measurement settings and $O(d^6)$ total copies are necessary for verification; (ii) systematic errors in state preparation or measurement are completely undetectable, which fundamentally undermines the quality of the estimate [194]. On the other hand, random errors, such as adding Gaussian noise to the measurement outcomes, are correctable with low overhead [208]. Additionally, heuristics have been studied to deal with the problem of systematic errors, under various assumptions about the process and the verifier [75, 30].

Gate set tomography (GST) [94, 159, 28, 27] is a direct attempt to mitigate drawback (ii), by explicitly parameterising systematic errors and by characterising

them alongside the process itself. Gate set tomography requires application of long strings of fundamental gate sequences, or "germs", in order to amplify and characterise coherent errors. Consequently, estimates of gate parameters are highly nonlinear functions of the measurement data, and extracting them with a high degree of accuracy is computationally challenging. GST does not mitigate the copy complexity of process tomography; GST of two trapped ion qubits in [169] required trials of 71 germs and 160 "fiducial" coherent errors, and a statistical analysis that required 12 hours of numerical optimisation and tens of gigabytes of memory.

Randomised benchmarking (RB) [76, 129, 146, 70] attempts to produce a coarser characterisation by testing; i.e. to output a quantity that is indicative, but not conclusive, about the operation of the process. The premise is to perform sequences of Clifford gates, chosen at random, that under perfect operation would leave the input state alone; when the operation is imperfect, the probability of projecting onto the input state at the output decays exponentially with the length of the sequence. One can then back out a single error parameter, the "RB number", as the exponent. Both the cost of post-processing [147] and total number of experiments [111] is, in principle, independent of the number of qubits; however, most techniques incur a costly overhead due to circuit compilation. The state of the art at time of writing is RB of five qubits [188]. It is also unclear what operational meaning to ascribe to the RB number, particularly in the case of gate dependent errors [187].

2.5 Comparison of verification strategies

To summarise, there is no panacea when verifying quantum states. An ideal verification protocol would: (i) maximise the information produced about the output state (for example, produce a full characterisation); (ii) make a minimal set of assumptions about the form of the output state; (iii) make a minimal set of assumptions about the abilities of the verifier; (iv) give the verifier the ability to make precise statements about statistical confidence given a finite number of trials; and (v) be efficient in terms of copy complexity. Regarding characterisation protocols that satisfy (i), practical tomography protocols only require simple measurements and formulaic post-processing but are prohibitively inefficient. Any saving in efficiency does not come for free; the verifier must either make some assumptions about the output state, as in constrained tomography, or empower the verifier to perform more complicated (and often physically implausible) measurements, as in sample-optimal tomography. Relaxing requirement (i), for example by relying on property testing protocols or direct fidelity estimation, is a plausible path to gains in efficiency. However, these protocols may still either rely on a verifier with access to non-local or collective measurements, or are prohibitively inefficient in practice.

To this end, we will seek out verification protocols in Chapter 3 that eschew a full characterisation of the output state, but are both efficient in copy complexity and rely on measurements that are typically available to a reasonable verifier. In Table 2.3 below, we give a comparison of previously established strategies for verifying a state $\rho \in \mathcal{D}(\mathbb{C}^{2^N})$ and the protocol in Chapter 3.

Strategy	**Copy complexity**	**States**	**Local**	**Single-copy**	**Non-adapt.**
"Pauli" tomography (§ 2.3.1)	$O\left(\frac{2^{4N}}{\epsilon_T^2}\log\frac{1}{\delta}\right)$	Arb.	✓	✓	✓
Single-copy tomography (§ 2.3.3)	$O\left(\frac{2^{3N}}{\epsilon_T^2}\log\frac{1}{\delta}\right)$	Arb.	✗	✓	✓
Sample-optimal tomography (§ 2.3.3)	$O\left(\frac{2^{2N}}{\epsilon_T^2}\log\frac{1}{\delta}\right)$	Arb.	✗	✗	✓
Shadow tomography (§ 2.1.5)	$\tilde{O}\left(\frac{N}{\epsilon^5}\log\frac{1}{\delta}\right)$	Arb.	✗	✗	✗
Compressed sensing (§ 2.3.2)	$O\left(\frac{2^{2N}}{\epsilon_T^2}\log\frac{1}{\delta}\right)$	Low rank	✓	✓	✓
Direct fidelity estimation (§ 2.1.5)	$O\left(\frac{2^{N}}{\epsilon^2}\log\frac{1}{\delta}\right)$	Pure	✓	✓	✓
Quantum state certification (§ 2.1.1)	$O\left(\frac{2^{N}}{\epsilon}\log\frac{1}{\delta}\right)$	Arb.	✗	✗	✓
Equality testing (§ 2.1.1)	$O\left(\frac{1}{\epsilon}\log\frac{1}{\delta}\right)$	Pure	✗	✓	✓
State verification (Chapter 3)	$O\left(\frac{1}{\epsilon}\log\frac{1}{\delta}\right)$	Stabil. (& 2-qubit)	✓	✓	✓

Table 2.3: A comparison of a selection of state verification strategies discussed in this chapter, and their copy complexities. "Arb." denotes arbitrary mixed states of N qubits, and "stabil." denotes stabilizer states of N qubits. For direct fidelity estimation, the target state is required to be pure but the tested state is not. Likewise, in compressed sensing the target state is required to be low rank but the tested state is only required to be close to a state of low rank.

CHAPTER 3

OPTIMAL VERIFICATION OF QUANTUM STATES WITH LOCAL MEASUREMENTS

3.1 Introduction

The goal of this chapter is to derive the optimal local verification strategy for commonly used entangled states and to compare its performance to bounds for competing protocols. In particular, we will compare performance with the bounds for precision-guaranteed, non-adaptive quantum state tomography by Sugiyama, Turner and Murao in [213], and the fidelity estimation protocol by Flammia and Yiu in [84]. Specifically, we will demonstrate non-adaptive verification strategies for arbitrary two-qubit states and stabilizer states of N qubits that are constructed from local measurements, and require quadratically fewer copies to verify to within a given fidelity than for these previous protocols. Moreover, we will see that the restriction to a verifier that is only allowed local measurements incurs only a constant factor penalty over the best non-local strategy, even if collective and adaptive measurements are allowed.

The breakdown of this chapter is as follows: in § 3.2, we outline the premise of verification protocols, and the types of protocols that we consider; we also introduce the statistical tools that will be necessary to analyse their performance in the context of asymmetric hypothesis tests. In § 3.3, we use this framework to treat the simplest case of verification of a Bell state. This result is extended to verification of arbitrary two qubit states in § 3.4, and then to verification of stabilizer states in § 3.5. We discuss a particular practical relaxation of the verification premise in § 3.6, before concluding with potential approaches to arbitrary state verification in § 3.7.

3.2 Quantum state verification protocols

3.2.1 Premise

Colloquially, a quantum state verification protocol is a procedure for gaining confidence that the output of some device is a particular state over any other. However, for any scheme involving measurements on a finite number of copies of the output state, one can always find an alternative state within some sufficiently small distance that is guaranteed to fool the verifier. Furthermore, the outcomes of measurements are, in general, probabilistic and a verification protocol collects a finite amount of data; and so any statement about verification can only be made up to some finite statistical confidence. The only meaningful statement to make in this context is the statistical inference that the state output from a device sits within a ball of a certain small radius (given some metric) of the correct state, with some statistical confidence. Thus the outcome of a state verification protocol is a statement like: "the device outputs copies of a state that has 99% fidelity with the target, with 90% probability". Note that this is different to the setting of state tomography; a verification protocol answers the question: *"Is the state* $|\psi\rangle$*?"* rather than the more involved tomographic question: *"Which state do I have?"*. Hence, unlike tomography, a verification protocol may give almost no information about the true state if the protocol fails.

We first proceed by setting up a formal framework for general state verification protocols. We assume that we have access to a device $\mathcal{D}$ that is supposed to produce copies of a state $|\psi\rangle$. However, $\mathcal{D}$ might not work correctly, and may actually produce (potentially mixed) states $\sigma_1, \sigma_2, \ldots$ such that σ_i might not be equal to $|\psi\rangle\langle\psi|$. In order to distinguish this from the case where the device works correctly by making a reasonable number of uses of $\mathcal{D}$, we need to have a promise that these states σ_i are sufficiently far from $|\psi\rangle$. So we are thus led to the following formulation of our verification task:

Distinguish between the following two cases:

(a). **(Good)** $\sigma_i = |\psi\rangle\langle\psi|$ for all i;

(b). **(Bad)** For some fixed ϵ, $F(|\psi\rangle, \sigma_i) := \langle\psi|\sigma_i|\psi\rangle \leq 1-\epsilon$ for all i.

Given a verifier with access to a set of available measurements $\mathcal{S}$, the protocols we consider for completing this task are of the following form:

Protocol Quantum state verification

1: **for** $i = 1$ to n **do**
2: Two-outcome measurement $M_i \in \mathcal{S}$ on σ_i, where M_i's outcomes are associated with "pass" and "fail"
3: **if** "fail" is returned **then**
4: Output "reject"
5: Output "accept"

We impose the conditions that in the good case, the protocol accepts with certainty, whereas in the bad case, the protocol accepts with probability at most δ; we call $1-\delta$ the *statistical power* of the protocol. We then aim to find a protocol that minimises n for a given choice of $|\psi\rangle$, ϵ and $\mathcal{S}$, such that these constraints are satisfied. Insisting that the protocol accepts in the good case with certainty implies that all measurements in $\mathcal{S}$ are guaranteed to pass in this case. This is a desirable property in itself, but one could consider more general non-adaptive protocols where measurements do not output "pass" with certainty on $|\psi\rangle$, and the protocol determines whether to accept based on an estimator constructed from the relative frequency of "pass" and "fail" outcomes across all n copies. We show in § 3.2.3 that this class of protocols has quadratically worse scaling in ϵ than protocols where each measurement passes with certainty on $|\psi\rangle$.

Additionally, protocols of the type we consider satisfy some useful operational properties:

A. *Non-adaptivity*. The strategy is fixed from the outset and depends only on the mathematical description of $|\psi\rangle$, rather than the choices of any prior measurements or their measurement outcomes.

B. *Future-proofing*. The strategy is independent of the infidelity ϵ, and gives a viable strategy for any choice of ϵ. Thus an experimentalist is able to arbitrarily decrease the infidelity ϵ within which verification succeeds by simply taking more total measurements following the strategy prescription, rather than modifying the prescription itself. The experimentalist is free to choose an arbitrary $\epsilon > 0$ and be guaranteed that the strategy still works in verifying $|\psi\rangle$.

We can make the following observations about this framework:

1. Given no restrictions on M_i, the optimal protocol is simply for each measurement to project onto $|\psi\rangle$. In fact, this remains optimal even over the class of more general protocols making use of adaptivity or collective measurements. One can see this as follows: if a two-outcome measurement M

(corresponding to the whole protocol) is described by measurement operators P (accept) and $\mathbb{1} - P$ (reject), then if M accepts $|\psi\rangle^{\otimes n}$ with certainty, we must have $P = |\psi\rangle\langle\psi|^{\otimes n} + P'$ for some residual positive semidefinite operator P'. Then replacing P with $|\psi\rangle\langle\psi|^{\otimes n}$ gives at least as good a protocol, as the probability of accepting $|\psi\rangle$ remains 1, while the probability of accepting other states cannot increase.

The probability of acceptance in the bad case after n trials is then at most $(1-\epsilon)^n$, so it is sufficient to take

$$n \geq \frac{\ln \delta^{-1}}{\ln((1-\epsilon)^{-1})} \approx \epsilon^{-1} \ln \delta^{-1} \tag{3.1}$$

to achieve statistical power $1-\delta$. This will be the yardstick against which we will compare our more restricted protocols below.

2. We assume that the states σ_i are independently and adversarially chosen. This implies that if (as we will consider below) $\mathcal{S}$ contains only projective measurements and does not contain the measurement projecting onto $|\psi\rangle\langle\psi|$, it is necessary to choose the measurement M_i at random from $\mathcal{S}$ and unknown to the adversary. Otherwise, we could be fooled with certainty by the adversary choosing σ_i to have support only in the "pass" eigenspace of M_i for each copy i.

3. We can be explicit about the optimisation needed to derive the optimal protocol in this adversarial setting. As protocols of the above form reject whenever a measurement fails, the adversary's goal at the i^{th} step is to maximise the probability that the measurement M_i at that step passes on σ_i. If the j^{th} measurement setting, M^j, is picked from $\mathcal{S}$ at step i with probability μ_j^i, the largest possible overall probability of passing for copy i is

$$\Pr[\text{Pass on copy } i] = \max_{\sigma_i, \langle\psi|\sigma_i|\psi\rangle \leq 1-\epsilon} \sum_j \mu_j^i \operatorname{tr}(P_j \sigma_i), \tag{3.2}$$

where we denote the corresponding "pass" projectors P_j. We can write $\Omega_i = \sum_j \mu_j^i P_j$, and then

$$\Pr[\text{Pass on copy } i] = \max_{\sigma, \langle\psi|\sigma|\psi\rangle \leq 1-\epsilon} \operatorname{tr}(\Omega_i \sigma). \tag{3.3}$$

As the verifier, we wish to minimise this expression over all Ω_i, so we end up with a final expression that does not depend on i. This leads us to infer that optimal protocols of this form can be assumed to be non-adaptive in two senses: they do not depend on the outcome of previous measurements (which is clear, as the protocol rejects if it ever sees a "fail" outcome); and they also do not depend on the measurement choices made previously.

Therefore, in order to find an optimal verification protocol, our task is to determine

$$\min_{\Omega} \max_{\sigma, \langle\psi|\sigma|\psi\rangle \leq 1-\epsilon} \mathrm{tr}(\Omega\sigma), \tag{3.4}$$

where Ω is an operator of the form $\Omega = \sum_j \mu_j P_j$ for $P_j \in \mathcal{S}$ and some probability μ_j. We call such operators *strategies*. If $\mathcal{S}$ contained all measurement operators (or even all projectors), Ω would be an arbitrary operator satisfying $0 \leq \Omega \leq I$. However, this notion becomes nontrivial when one considers physically-motivated restrictions on $\mathcal{S}$.

4. In a non-adversarial scenario, it may be acceptable to fix the measurements in Ω in advance, with appropriate frequencies μ_j. Then, given n, a strategy $\Omega = \sum_j \mu_j P_j$ corresponds to a protocol where for each j we deterministically make $\mu_j n$ measurements $\{P_j, \mathbb{1} - P_j\}$. For large n, and fixed $\sigma_i = \sigma$, this will achieve similar performance to the above protocol.

5. More complicated protocols with adaptive or collective measurements, or measurements with more than two outcomes, cannot markedly improve on the strategies derived here. We do not treat these more general strategies explicitly, but note that the protocols we will describe based on local projective measurements already achieve the globally optimal bound in Eq. 3.1 up to constant factors, so any gain from these more complex approaches would be minor.

We have shown that, given no constraints on the verifier's measurement prescription, the optimal strategy is to just project on to $|\psi\rangle$; however, in general the projector $|\psi\rangle\langle\psi|$ will be non-local, which has the disadvantage of being harder to implement experimentally. This is particularly problematic in linear quantum optics, for example, where deterministic, unambiguous discrimination of a complete set of Bell states is impossible [222, 40, 78]. Thus, for each copy there is a fixed probability of the measurement returning a "null" outcome; hence, regardless of the optimality of the verification strategy, merely the probability of its successful operation decreases exponentially with the number of measurements. Alternatively, one may consider a verification protocol in a distributed setting, where parties are sent some part of a quantum state and must measure locally before classically conferring their outcomes with other parties. As such, we seek optimal measurement strategies that satisfy some natural properties that make them both physically realisable and useful to a real-world verifier. We impose the following properties:

1. *Locality*. $\mathcal{S}$ contains only measurements corresponding to local observables, acting on a single copy of the output state.

2. *Projective measurement*. $\mathcal{S}$ contains only binary-outcome, projective measurements, rather than more elaborate POVMs.

3. *Trust*. The physical operation of each measurement device is faithful to its mathematical description; it behaves as expected, without experimental error.

Thus for multipartite states we only consider strategies where each party locally performs a projective measurement on a single copy, and the parties accept or reject based on their collective measurement outcomes. We also highlight the trust requirement to distinguish from the self-testing protocols outlined in § 2.1.2.

3.2.2 The Chernoff-Stein lemma

In the quantum state verification scenario where we only have access to single-copy measurements, even if the systems under scrutiny are quantum mechanical the data we collect are ostensibly classical, and so any test we construct falls under the jurisdiction of classical information theory. The premise we have outlined in § 3.2.1 is of a *binary hypothesis test* (see § 2.1.6 for an introduction). The goal of this section is to give a pedagogical introduction to results that scrutinise the ultimate performance of binary hypothesis tests, culminating in the *Chernoff-Stein lemma*. This will be the jumping-off point from which we derive the results on verification of quantum states in the following sections. The following discussion closely follows [68], § 11.8; we refer the reader there for a more thorough treatment. Conversely, any reader that is either familiar with, or disinterested in, classical estimation theory can safely skip this section and return to quantum protocols in § 3.2.3.

We will consider an experimental setup described by a set of n trials, each governed by an independent and identically-distributed random variable X_i. An experiment $X_1, X_2 \dots X_n$ will have a set of outcomes, which we denote $x_1, x_2 \dots x_n$, or alternatively by the vector $\mathbf{x} = (x_1, x_2 \dots x_n)$. The set of all possible measurement outcomes, i.e. all possible choices of $\mathbf{x}$, we denote as $\mathcal{X}$. Our null and alternative hypotheses, H_0 and H_1, will simply be that the outcomes are drawn from a probability distribution P_0 or P_1, respectively. We assume that we are promised that one of these two cases is true, and we must decide which. For any test that we construct, there will be some set of outcomes $\mathbf{x}$ where we accept H_0, and some set of outcomes where we accept H_1. We'll call these sets $A_0 \subseteq \mathcal{X}$ and $A_1 \subseteq \mathcal{X}$, respectively. Given the promise that either H_0 or H_1 is true, we have that $A_1 = \bar{A}_0$ (where the overbar denotes the set complement).

A quantity that will routinely feature in this context is the *relative entropy*, or *Kullback-Leibler divergence*, between two distributions; which we denote $D(\cdot\|\cdot)$. As a reminder, for the case of discrete probability distributions P_0 and P_1, it is defined as

$$D(P_0\|P_1) = \sum_{\mathbf{x}\in\mathcal{X}} P_0(\mathbf{x}) \log \frac{P_0(\mathbf{x})}{P_1(\mathbf{x})}. \tag{3.5}$$

While it is commonly asserted that the relative entropy is a measure of dissimilarity between two probability distributions, some care must be taken as it is not a true metric. In particular, it is not guaranteed that D is symmetric; nor is it guaranteed to satisfy the triangle inequality. The simplest colloquial description of the relative entropy is that it is a quantifier of surprise; in a Bayesian sense, $D(P_0\|P_1)$ quantifies the amount of information gained in updating from a prior P_0 to a posterior P_1. In hypothesis testing, the relative entropy appears as the limit of the likelihood ratio of two hypotheses:

Lemma 10. *If $X_1 \ldots X_n$ is a sequence of iid random variables drawn from P_0, and P_1 is any other distribution, then*

$$\frac{1}{n} \log \frac{P_0(X_1 \ldots X_n)}{P_1(X_1 \ldots X_n)} \to D(P_0\|P_1), \tag{3.6}$$

where $\to$ denotes convergence in probability in the limit of large n.

Proof. Rearranging the lefthand side assuming that the trials are independent and identically distributed gives

$$\frac{1}{n} \log \frac{P_0(X_1 \ldots X_n)}{P_1(X_0 \ldots X_n)} = \frac{1}{n} \log \frac{\prod_i P_0(X_i)}{\prod_i P_1(X_i)} = \frac{1}{n} \sum_i \log \frac{P_0(X_i)}{P_1(X_i)}. \tag{3.7}$$

Given the weak law of large numbers, this expression converges in probability to $\mathbb{E}_{P_0}[\log(P_0(X)/P_1(X))] = D(P_0\|P_1)$. □

While Lemma 10 is a precise statement about the asymptotics of the log-likelihood ratio, it is not the case for a fixed and finite n that the log-likelihood ratio and the relative entropy are guaranteed to be close. However, we will make use of a subset of sequences of outcomes, that occurs with high probability and where the log-likelihood ratio and the relative entropy are guaranteed to be close. If a sequence of outcomes $\mathbf{x} \in \mathcal{X}$ satisfies

$$D(P_0\|P_1) - \epsilon \leq \frac{1}{n} \log \frac{P_0(\mathbf{x})}{P_1(\mathbf{x})} \leq D(P_0\|P_1) + \epsilon \tag{3.8}$$

for a fixed and finite n and ϵ then we call the sequence of outcomes $\mathbf{x}$ "typical"; and we use notation $X_n^\epsilon(P_0\|P_1)$ (or the shorthand X_n^ϵ when the distributions are clear from context) to denote the set of all sequences of outcomes that are typical for a particular n and ϵ. Typical sequences satisfy some nice operational properties:

Lemma 11. *For a typical* $\mathbf{x}$ *(i.e.* $\mathbf{x} \in X_n^\epsilon(P_0\|P_1)$*):*

1. *The probability that a sequence is drawn from the alternative distribution* P_1 *is bounded:*
$$P_0(\mathbf{x})2^{-n[D(P_0\|P_1)+\epsilon]} \leq P_1(\mathbf{x}) \leq P_0(\mathbf{x})2^{-n[D(P_0\|P_1)-\epsilon]}. \tag{3.9}$$

2. *Let* $P_0(X_n^\epsilon(P_0\|P_1)) = \sum_{\mathbf{x}\in X_n^\epsilon} P_0(\mathbf{x})$*. Then the expected probability that we successfully guess that a typical sequence is drawn from distribution* P_0 *is*
$$P_0(X_n^\epsilon(P_0\|P_1)) > 1-\epsilon, \tag{3.10}$$
for large enough n*.*

3. *Let* $P_1(X_n^\epsilon(P_0\|P_1)) = \sum_{\mathbf{x}\in X_n^\epsilon} P_1(\mathbf{x})$*. Then the expected probability that we guess that a typical sequence is drawn from the alternative distribution* P_1 *is*
$$P_1(X_n^\epsilon(P_0\|P_1)) < 2^{-n[D(P_0\|P_1)-\epsilon]}. \tag{3.11}$$

Proof. Eq. 3.9 follows directly from rearranging the definition of a typical $\mathbf{x}$, and Eq. 3.10 is just a restatement of Lemma 10. Eq. 3.11 follows from the following:
$$P_1(X_n^\epsilon(P_0\|P_1)) = \sum_{\mathbf{x}\in X_n^\epsilon} P_1(\mathbf{x}) \leq \sum_{\mathbf{x}\in X_n^\epsilon} P_0(\mathbf{x})2^{-n[D(P_0\|P_1)-\epsilon]}, \tag{3.12}$$
where the inequality is from Eq. 3.9. Then since $\sum_{\mathbf{x}} P_0(\mathbf{x} \in X) \leq 1$ for all possible subsets X, Eq. 3.11 immediately follows. □

Suppose that we run our experiment, and get a sequence of outcomes drawn from some subset $S \subset \mathcal{X}$. We are interested in an asymmetric hypothesis test, where one type of error probability is fixed; say $P_0(S) > 1-\epsilon$. Then we can bound the other probability of error using the properties of typical sequences:

Lemma 12. *For* $S \subset \mathcal{X}$ *such that* $P_0(S) > 1-\epsilon$*, then for any other distribution* P_1*,*
$$P_1(S) > (1-2\epsilon)2^{-n[D(P_0\|P_1)+\epsilon]}. \tag{3.13}$$

Proof. We start by noting that $P_1(S) \geq P_1(X_n^\epsilon \cap S)$, for any set of typical sequences X_n^ϵ. So, if we also make use of the lower bound on $P_1(\mathbf{x})$ in Eq. 3.9, we have that
$$\begin{aligned} P_1(S) \geq P_1(X_n^\epsilon \cap S) &= \sum_{\mathbf{x}\in S\cap X_n^\epsilon} P_1(\mathbf{x}) \\ &\geq \sum_{\mathbf{x}\in X_n^\epsilon \cap S} P_0(\mathbf{x})2^{-n[D(P_0\|P_1)+\epsilon]} \\ &= 2^{-n[D(P_0\|P_1)+\epsilon]} P_0(X_n^\epsilon \cap S). \end{aligned} \tag{3.14}$$

Now, we know from the initial assumption about S that $P_0(S) > 1-\epsilon$, and we know from Eq. 3.10 that $P_0(X_n^\epsilon) > 1-\epsilon$. Additionally, $P_0(S \cup X_n^\epsilon) \leq 1$, as $P_0(\cdot)$ denotes a probability. Then using the inclusion-exclusion principle,

$$P_0(X_n^\epsilon \cap S) = P_0(X_n^\epsilon) + P_0(S) - P_0(X_n^\epsilon \cup S) \geq 1-\epsilon+1-\epsilon-1 = 1-2\epsilon. \tag{3.15}$$

Combining Eqs. 3.15 and 3.14 gives the desired result. □

These ingredients are enough to derive the optimal rate at which one can perform an asymmetric hypothesis test. In general, given a fixed Type I error, the optimum asymptotic rate at which the Type II error can be minimised in an asymmetric hypothesis test is given by the *Chernoff-Stein lemma*:

Theorem 13 (Cover and Thomas [68], Theorem 11.8.3.)**.** *Let $X_1 \ldots X_n$ be drawn iid from a probability mass function Q. Then consider the hypothesis test between alternatives H_0: $Q = P_0$ and H_1: $Q = P_1$. Let $A_0 \subset \mathcal{X}$ be the subset of outcome sequences where we conclude that the null hypothesis is correct. Denote Type I and Type II errors after n samples as α_n^* and β_n^*, respectively. Then for some constraint parameter $0 < \epsilon < \frac{1}{2}$, define the minimal Type II error given a constrained Type I error as*

$$\delta_n^\epsilon = \min_{\substack{A_0 \\ \alpha_n^* < \epsilon}} \beta_n^*. \tag{3.16}$$

Then asymptotically, the rate at which this error diminishes given increasing trials is

$$\lim_{n\to\infty} \frac{1}{n} \ln \delta_n^\epsilon = -D(P_0 \,\|P_1), \tag{3.17}$$

where $D(P_0\|P_1)$ is the relative entropy between probability distributions P_0 and P_1.

Proof. We must show two things for this theorem to be correct: (i) that there exists a particular choice of A_0 that satisfies Eq. 3.17; and (ii) that no other choice can give a steeper decrease. For (i), take $A_0 = X_n^\epsilon$. Then we are guaranteed that $\alpha_n^* < \epsilon$, by Eq. 3.10, and we are guaranteed that the rate is at most $-(D(P_0\|P_1)-\epsilon)$, by Eq. 3.11. For (ii), we know from Lemma 3.13 that $P_1(S)$ is bounded from below; so we can rearrange and take the limit $n \to \infty$ to yield

$$\lim_{n\to\infty} \frac{1}{n} \log P_1(S) > -(D(P_0\|P_1)+\epsilon) + \lim_{n\to\infty} \frac{1}{n} \log(1-2\epsilon) = -(D(P_0\|P_1)+\epsilon). \tag{3.18}$$

This is strictly smaller in magnitude than the rate for X_n^ϵ, and so this choice is asymptotically optimal in the rate at which the Type II error decreases for fixed Type I error. □

For clarity, we drop the sub- and superscript $\delta_n^\epsilon \to \delta$. After rearranging the expression for the optimal asymptotic Type II error given by the Chernoff-Stein lemma, we can asymptotically achieve a test to discriminate P_0 and P_1 with statistical power $1-\delta$ by taking a number of measurements

$$n > \frac{1}{D(P_0\|P_1)} \ln \frac{1}{\delta}. \tag{3.19}$$

Moreover, this bound is tight in that it gives the correct asymptotic relationship between n, D and δ; generically δ can be lower bounded ([68], p666) such that

$$\frac{e^{-Dn}}{n+1} \leq \delta \leq e^{-Dn}. \tag{3.20}$$

3.2.3 State verification and hypothesis testing

The state verification scenario that we have outlined above is a binary hypothesis test between the case labelled "Good", where each state output from the experiment is $|\psi\rangle$, and the case labelled "Bad" where each state output is far from $|\psi\rangle$. We have asserted the following proposition:

Proposition 14. *Any strategy Ω that: (a) accepts $|\psi\rangle$ with certainty, $p := \mathrm{tr}(\Omega|\psi\rangle\langle\psi|) = 1$; and (b) does not accept σ with certainty ($\mathrm{tr}(\Omega\sigma) := p - \Delta_\epsilon$, for $\Delta_\epsilon > 0$) requires asymptotically fewer measurements in infidelity ϵ to distinguish these states to within a fixed Type II error than the best protocol based on a strategy Ω' where $\mathrm{tr}(\Omega'|\psi\rangle\langle\psi|) < 1$.*

Proposition 14 states that, in a framework where we attempt to verify $|\psi\rangle$ by repeating two-outcome measurements picked from some set, asymptotically it is always beneficial to use measurements that accept $|\psi\rangle$ with certainty. In this case, each measurement is a Bernoulli trial with some acceptance probability. An example of a protocol which would *not* satisfy this property would be estimating the probability of violating a Bell inequality for a maximally entangled two qubit state.

In this case, we have shown the verifier's choice of Ω is fixed, independent of the trial, and that in the adversarial scenario the fooling state σ is also fixed. As such, each measurement is just a Bernoulli trial with some fixed acceptance probability. As demonstrated in § 3.2.2, the best asymptotic rate at which one could expect the probability of error to decrease in this hypothesis test is given by the relative entropy, as dictated by the Chernoff-Stein lemma (Thm. 13). In particular, the number of copies needed is

$$n = \frac{1}{D(p\|p-\Delta_\epsilon)} \log \frac{1}{\delta}. \tag{3.21}$$

The relative entropy typically takes a pair of probability distributions as arguments, but given that each hypothesis is concerned only with a single Bernoulli-distributed random variable uniquely specified by a pair of real parameters, we will use the shorthand $D(a\|b)$ for Bernoulli parameters a and b. In this case the relative entropy can be expanded as

$$D(a\|b) = a\ln\frac{a}{b} + (1-a)\ln\frac{1-a}{1-b}. \tag{3.22}$$

Two important limiting cases of this expression have relevance here. Firstly, if the gap between p and $p-\Delta_\epsilon$ is small, then Taylor expanding the expression for the relative entropy for small Δ_ϵ gives that it is sufficient to take

$$n \geq \frac{2p(1-p)}{\Delta_\epsilon^2}\ln\frac{1}{\delta}. \tag{3.23}$$

Secondly, in the limit where $p \to 1$, using that $\lim_{p\to 1^-}(1-p)\ln(1-p) = 0$, the relative entropy becomes

$$\lim_{p\to 1^-} D(p\|p-\Delta_\epsilon) = \ln\frac{1}{1-\Delta_\epsilon}; \tag{3.24}$$

and so the number of samples scales like

$$n \geq \frac{-1}{\log(1-\Delta_\epsilon)}\ln\frac{1}{\delta} \approx \frac{1}{\Delta_\epsilon}\ln\frac{1}{\delta}. \tag{3.25}$$

These are the limiting cases of the scaling of n with Δ_ϵ, the gap between the acceptance probabilities for $|\psi\rangle$ and a state ϵ away. In the worst case, n scales quadratically in Δ_ϵ^{-1}; however, for any strategy where one of the Bernoulli hypotheses to be tested is accepted with certainty, only a total number of measurements linear in Δ_ϵ^{-1} are required. Thus asymptotically, a strategy where $p = 1$ is always favourable (i.e. gives a quadratic improvement in scaling with Δ_ϵ) for any $\Delta_\epsilon > 0$. Hence any strategy that is constructed from measurements that accept $|\psi\rangle$ with certainty requires quadratically fewer copies to verify to within a fixed Δ_ϵ, in general.

3.2.4 Verification strategy optimisation

In this section, we simplify the form of the optimisation in Eq. 3.4 using the strategy requirements outlined above. We start by making the following useful observation:

Lemma 15. *We can assume without loss of generality that, in Eq. 3.4, σ is pure.*

Proof. Assume the adversary chooses a fixed density matrix σ, which is globally optimal: it forces the verifier to accept σ with the greatest probability among states σ such that $\langle\psi|\sigma|\psi\rangle := r \leq 1-\epsilon$. The probability of accepting this σ given strategy Ω is then

$$\Pr[\text{Accept } \sigma] = \mathrm{tr}(\Omega\sigma). \tag{3.26}$$

We have asserted that Ω accepts $|\psi\rangle$ with certainty: $\langle\psi|\Omega|\psi\rangle = 1$. However, for this to be the case Ω must have $|\psi\rangle$ as an eigenstate with eigenvalue 1; thus we can write

$$\Omega = |\psi\rangle\langle\psi| + \sum_j c_j |\psi_j^\perp\rangle\langle\psi_j^\perp| \tag{3.27}$$

where the states $\{|\psi_j^\perp\rangle\}$ are a set of mutually orthogonal states orthogonal to $|\psi\rangle$. Then

$$\Pr[\text{Accept } \sigma] = \langle\psi|\sigma|\psi\rangle + \sum_j c_j \langle\psi_j^\perp|\sigma|\psi_j^\perp\rangle \tag{3.28}$$

$$= r + \sum_j c_j \langle\psi_j^\perp|\sigma|\psi_j^\perp\rangle. \tag{3.29}$$

We can write

$$\sigma = a|\psi\rangle\langle\psi| + b\sigma^\perp + c|\psi\rangle\langle\Phi^\perp| + c^*|\Phi^\perp\rangle\langle\psi|, \tag{3.30}$$

where $\sigma^\perp$ is a density matrix entirely supported in the subspace spanned by the states $|\psi_j^\perp\rangle$, and $|\Phi^\perp\rangle$ is a vector in the subspace spanned by $|\psi_j^\perp\rangle$. We know that $a = r$ as $\langle\psi|\sigma|\psi\rangle = r$, and $b = 1 - r$ as $\mathrm{tr}(\sigma) = 1$. Then,

$$\Pr[\text{Accept } \sigma] = r + \sum_j c_j \langle\psi_j^\perp|\left[r|\psi\rangle\langle\psi| + (1-r)\sigma^\perp + c|\psi\rangle\langle\Phi^\perp| + c^*|\Phi^\perp\rangle\langle\psi|\right]|\psi_j^\perp\rangle \tag{3.31}$$

$$= r + (1-r)\sum_j c_j \langle\psi_j^\perp|\sigma^\perp|\psi_j^\perp\rangle. \tag{3.32}$$

Since the summation in Eq. 3.32 is an average over expectation values, the adversary cannot do better than picking the single largest expectation value in the summation; they must take $\sigma^\perp$ to be the state $|\psi_{max}^\perp\rangle\langle\psi_{max}^\perp|$, where $|\psi_{max}^\perp\rangle$ is the state $|\psi_j^\perp\rangle$ in the spectral decomposition of Ω with largest eigenvalue, c_{max}. Thus

$$\max_\sigma \mathrm{tr}(\Omega\sigma) = r + (1-r)c_{max}, \tag{3.33}$$

which is achieved by any density matrix of the form

$$\sigma = r|\psi\rangle\langle\psi| + (1-r)|\psi_{max}^\perp\rangle\langle\psi_{max}^\perp| + c|\psi\rangle\langle\Phi^\perp| + c^*|\Phi^\perp\rangle\langle\psi|. \tag{3.34}$$

Note that the pure state $\sigma = |\phi\rangle\langle\phi|$ for $|\phi\rangle = \sqrt{r}|\psi\rangle + \sqrt{1-r}|\psi_{max}^\perp\rangle$ is of this form, and so we can assume that the adversary makes this choice. □

Given that the state σ can be taken to be pure and that the fidelity $F(|\psi\rangle, \sigma) \leq 1-\epsilon$, we write $\sigma = |\psi_{\bar\epsilon}\rangle\langle\psi_{\bar\epsilon}|$, where $|\psi_{\bar\epsilon}\rangle := \sqrt{1-\bar\epsilon}|\psi\rangle + \sqrt{\bar\epsilon}|\psi^\perp\rangle$ and $\langle\psi|\psi^\perp\rangle = 0$, for some $\bar\epsilon \geq \epsilon$ chosen by the adversary, to be optimised later. Denote

$$\min_\Omega \max_{\substack{\sigma \\ \langle\psi|\sigma|\psi\rangle \leq 1-\epsilon}} \mathrm{tr}(\Omega\sigma) := 1 - \Delta_\epsilon. \tag{3.35}$$

Then the optimisation problem becomes to determine Δ_ϵ, where

$$\Delta_\epsilon = \max_{\Omega} \min_{|\psi^\perp\rangle, \bar{\epsilon} \geq \epsilon} \bar{\epsilon}(1 - \langle\psi^\perp|\Omega|\psi^\perp\rangle) - 2\sqrt{\bar{\epsilon}(1-\bar{\epsilon})}\mathrm{Re}(\langle\psi|\Omega|\psi^\perp\rangle) \tag{3.36}$$

and $\Omega|\psi\rangle = |\psi\rangle$.

This expression can be simplified given that $\Omega|\psi\rangle = |\psi\rangle$. In particular, we then know that $\langle\psi^\perp|\Omega|\psi\rangle = 0$ for any choice of orthogonal state $|\psi^\perp\rangle$. Thus the term $\sqrt{\bar{\epsilon}(1-\bar{\epsilon})}\mathrm{Re}(\langle\psi|\Omega|\psi^\perp\rangle)$ automatically vanishes. We are then left with the optimisation

$$\Delta_\epsilon = \max_{\Omega} \min_{|\psi^\perp\rangle, \bar{\epsilon} \geq \epsilon} \bar{\epsilon}(1 - \langle\psi^\perp|\Omega|\psi^\perp\rangle), \tag{3.37}$$

where $\Omega|\psi\rangle = |\psi\rangle$.

As for the optimisation of $\bar{\epsilon}$, note that it is the goal of the adversary to make Δ_ϵ as small as possible; and so they are obliged to set $\bar{\epsilon} = \epsilon$. Then the optimisation becomes

$$\Delta_\epsilon = \epsilon \max_{\Omega} \min_{|\psi^\perp\rangle} (1 - \langle\psi^\perp|\Omega|\psi^\perp\rangle), \tag{3.38}$$

where $\Omega|\psi\rangle = |\psi\rangle$.

Note that this expression implies that any Ω where $\Omega|\psi\rangle = |\psi\rangle$ automatically satisfies the *future-proofing* property: firstly that Ω is independent of ϵ, but also that the strategy must be viable for any choice of ϵ (i.e. there must not be a choice of ϵ where $\Delta_\epsilon = 0$). For an initial choice $\Delta_\epsilon > 0$, we have that $1 - \langle\psi^\perp|\Omega|\psi^\perp\rangle > 0$ and so $\Delta_{\epsilon'} > 0$ for any $0 < \epsilon' < \epsilon$. Thus the verifier is free to decrease ϵ arbitrarily without fear of the strategy failing. Note also that this condition may not be automatically guaranteed if the verifier chooses an Ω such that $\Omega|\psi\rangle \neq |\psi\rangle$.

Regarding the optimisation problem in Eq. 3.38, for an arbitrary state $|\psi\rangle$ on n qubits it is far from clear how to: (a) construct families of viable Ω (built from local projective measurements) that accept $|\psi\rangle$ with certainty; (b) to then solve this optimisation problem over those families of Ω. For the remainder of this work, we focus on states of particular experimental interest where we can solve the problem: arbitrary states of two qubits, and stabilizer states. To illustrate our approach, we start with the case of a Bell state before generalising to larger classes of states.

3.3 Warm-up: Bell state verification

Consider the case of verifying the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. If we maintain a strategy where all measurements accept $|\Phi^+\rangle$ with certainty, then it must be the case that $\Omega|\Phi^+\rangle = |\Phi^+\rangle$. The optimisation problem for the verifier-adversary pair is then given by Δ_ϵ:

$$\Delta_\epsilon = \max_{\Omega} \min_{\substack{\sigma \\ \langle\psi|\sigma|\psi\rangle \leq 1-\epsilon}} \mathrm{tr}[\Omega(|\Phi^+\rangle\langle\Phi^+| - \sigma)]. \tag{3.39}$$

However, we have shown in § 3.2.4 that it is never beneficial for the adversary to: (a) choose a non-pure σ; or (b) to pick a σ such that $\langle\psi|\sigma|\psi\rangle < 1-\epsilon$. Rewrite $\sigma = |\psi_\epsilon\rangle\langle\psi_\epsilon|$, where $|\psi_\epsilon\rangle = \sqrt{1-\epsilon}|\Phi^+\rangle + \sqrt{\epsilon}|\psi^\perp\rangle$ for some state $|\psi^\perp\rangle$ such that $\langle\Phi^+|\psi^\perp\rangle = 0$. Then,

$$\begin{aligned}\Delta_\epsilon = \max_{\Omega} \min_{|\psi^\perp\rangle} & \epsilon(\langle\Phi^+|\Omega|\Phi^+\rangle - \langle\psi^\perp|\Omega|\psi^\perp\rangle) \\ & - 2\sqrt{\epsilon(1-\epsilon)}\mathrm{Re}\langle\Phi^+|\Omega|\psi^\perp\rangle.\end{aligned} \tag{3.40}$$

Given that $\Omega|\Phi^+\rangle = |\Phi^+\rangle$, we can simplify by noting that $\langle\Phi^+|\Omega|\Phi^+\rangle = 1$ and $\langle\Phi^+|\Omega|\psi^\perp\rangle = 0$. Thus,

$$\begin{aligned}\Delta_\epsilon &= \max_{\Omega} \min_{|\psi^\perp\rangle} \epsilon(1 - \langle\psi^\perp|\Omega|\psi^\perp\rangle) \\ &= \epsilon(1 - \min_{\Omega} \max_{|\psi^\perp\rangle} \langle\psi^\perp|\Omega|\psi^\perp\rangle),\end{aligned} \tag{3.41}$$

where the verifier controls Ω and the adversary controls $|\psi^\perp\rangle$. Given that $|\Phi^+\rangle$ is itself an eigenstate of Ω, the worst-case scenario for the verifier is for the adversary to choose $|\psi^\perp\rangle$ as the eigenstate of Ω with the next largest eigenvalue. If we diagonalise Ω we can write $\Omega = |\Phi^+\rangle\langle\Phi^+| + \sum_{j=1}^{3} \nu_j |\psi_j^\perp\rangle\langle\psi_j^\perp|$, where $\langle\Phi^+|\psi_j^\perp\rangle = 0 \; \forall j$. The adversary picks the state $|\psi_{\max}^\perp\rangle$ with corresponding eigenvalue $\nu_{\max} = \max_j \nu_j$. Now, consider the trace of Ω: if $\mathrm{tr}(\Omega) < 2$ then the strategy must be a convex combination of local projectors, at least one of which is rank 1. However, the only rank 1 projector that satisfies $P^+|\Phi^+\rangle = |\Phi^+\rangle$ is $P^+ = |\Phi^+\rangle\langle\Phi^+|$, which is non-local; and therefore $\mathrm{tr}(\Omega) \geq 2$. Combining this with the expression for Ω above gives $\mathrm{tr}(\Omega) = 1 + \sum_j \nu_j \geq 2$. It is always beneficial to the verifier to saturate this inequality, as any extra weight on the subspace orthogonal to $|\Phi^+\rangle$ can only increase the chance of being fooled by the adversary. Thus the verifier is left with the optimisation

$$\min \nu_{\max} = \min \max_k \nu_k, \quad \sum_k \nu_k = 1. \tag{3.42}$$

This expression is optimised for $\nu_j = \frac{1}{3}, j = 1,2,3$. In this case, $\Omega = |\Phi^+\rangle\langle\Phi^+| + \frac{\mathbb{1}}{3}$. Then we can rewrite Ω as

$$\Omega = \frac{1}{3}(P_{XX}^+ + P_{-YY}^+ + P_{ZZ}^+), \tag{3.43}$$

where P^+_{XX} is the projector onto the positive eigensubspace of the tensor product of Pauli matrices XX (and likewise for $-YY$ and ZZ). The operational interpretation of this optimal strategy is then explicit: for each copy of the state, the verifier randomly chooses a measurement setting from the set $\{XX, -YY, ZZ\}$ all with probability $\frac{1}{3}$, and accepts only on receipt of outcome "+1" on all n measurements. Note that we could expand Ω differently, for example by conjugating each term in the above expression by any local operator that leaves $|\Phi^+\rangle$ alone; the decomposition above is only one of a family of optimal strategies. As for scaling, we know that $\Delta_\epsilon = \epsilon(1 - \nu_{\max}) = \frac{2\epsilon}{3}$, and the number of measurements needed to verify the Bell state $|\Phi^+\rangle$ is then $n_{opt} = \left[\ln\left(\frac{3}{3-2\epsilon}\right)\right]^{-1} \ln\frac{1}{\delta} \approx \frac{3}{2\epsilon}\ln\frac{1}{\delta}$. Note that this is only worse than the optimal non-local strategy by a factor of 1.5.

In comparison, consider instead verifying a Bell state by performing a CHSH test. Then even in the case of trusted measurements, the total number of measurements scales like $O\left(\frac{1}{\epsilon^2}\right)$ [212] (see § 2.1.3), which is quadratically worse than the case of measuring the stabilizers $\{XX, -YY, ZZ\}$. This suboptimal scaling is shared by the known bounds for non-adaptive quantum state tomography with single-copy measurements in [213] and fidelity estimation in [84]. See § 2.3.1 and [206, 82, 210] for further discussion of this scaling in tomography. Additionally, two-qubit tomography potentially requires five times as many measurement settings. We also note that a similar quadratic improvement was derived in adaptive quantum state tomography in [149], in the sample-optimal tomographic scheme in [100] and in the quantum state certification scheme in [11]; however, the schemes therein assume access to either non-local or collective measurements.

3.4 Verifying arbitrary states of two qubits

The goal is unchanged for other pure states of two qubits: we seek strategies that accept the target state with certainty, and hence achieve the asymptotic advantage outlined for Bell states above. It is not clear a priori that such a strategy exists for general states, in a way that is as straightforward as the previous construction. However, we show that for any two-qubit state not only does such a strategy exist, but we can optimise within the family of allowable strategies and give an analytic expression with optimal constant factors.

We first remark that we can restrict to states of the form $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$ without loss of generality, as any state is locally equivalent to a state of this form, for some θ. Specifically:

Lemma 16. *Given any two qubit state $|\psi\rangle$ with optimal strategy Ω_{opt}, a locally equivalent state $(U \otimes V)|\psi\rangle$ has optimal strategy $(U \otimes V)\Omega_{opt}(U \otimes V)^\dagger$.*

Proof. We must show that strategy $\Omega' = (U \otimes V)\Omega_{opt}(U \otimes V)^\dagger$ is both a valid strategy, and is optimal for verifying $|\psi'\rangle = (U \otimes V)|\psi\rangle$.
Validity: If $\Omega_{opt} = \sum_j \mu_j P_j$ is a convex combination of local projectors, then so is Ω':

$$\begin{aligned}\Omega' = (U \otimes V)\Omega(U \otimes V)^\dagger &= \sum_j \mu_j (U \otimes V)P_j(U \otimes V)^\dagger \\ &= \sum_j \mu_j P'_j. \end{aligned} \tag{3.44}$$

Also, if $\Omega_{opt}|\psi\rangle = |\psi\rangle$ then $\Omega'|\psi'\rangle = |\psi'\rangle$:

$$\begin{aligned}\Omega_{opt}|\psi\rangle = |\psi\rangle &\Rightarrow (U \otimes V)\Omega|\psi\rangle = p_{opt}(U \otimes V)|\psi\rangle \\ &\Rightarrow (U \otimes V)\Omega(U \otimes V)^\dagger(U \otimes V)|\psi\rangle = (U \otimes V)|\psi\rangle \\ &\Rightarrow \Omega'|\psi'\rangle = |\psi'\rangle. \end{aligned} \tag{3.45}$$

Optimality: The performance of a strategy is determined by the maximum probability of accepting an orthogonal state $|\psi^\perp\rangle$. For the strategy-state pairs $(\Omega_{opt}, |\psi\rangle)$ and $(\Omega', |\psi'\rangle)$, we denote this parameter q_{opt} and q', respectively. Then

$$q_{opt} = \max_{|\psi^\perp\rangle} \langle\psi^\perp|\Omega_{opt}|\psi^\perp\rangle = \max_{|\phi\rangle, \langle\psi|\phi\rangle = 0} \langle\phi|\Omega_{opt}|\phi\rangle \tag{3.46}$$

$$= \max_{(U\otimes V)|\phi\rangle, \langle\psi|(U\otimes V)^\dagger(U\otimes V)|\phi\rangle = 0} \langle\phi|(U \otimes V)^\dagger(U \otimes V)\Omega_{opt}(U \otimes V)^\dagger(U \otimes V)|\phi\rangle \tag{3.47}$$

$$= \max_{|\phi'\rangle, \langle\psi'|\phi'\rangle = 0} \langle\phi'|\Omega'|\phi'\rangle = q'. \tag{3.48}$$

So applying the same local rotation to the strategy and the state results in no change in the performance of the strategy. Thus the following simple proof by contradiction holds: assume that there is a better strategy for verifying $|\psi'\rangle$, denoted Ω''. But then the strategy $(U \otimes V)^\dagger\Omega''(U \otimes V)$ must have a better performance for verifying $|\psi\rangle$ than Ω_{opt}, which is a contradiction. Thus Ω' must be the optimal strategy for verifying $|\psi'\rangle$. □

We are close to being able to derive the optimal local strategy for states of two qubits. However, we first prove a useful (and at face value, a somewhat obvious) lemma - that no optimal strategy can contain the identity measurement (i.e. it is never beneficial for a verifier to always accept, regardless of the tested state). In the following discussion, we denote the projector $\Pi := \mathbb{1} - |\psi\rangle\langle\psi|$. For a strategy Ω where $\Omega|\psi\rangle = |\psi\rangle$, the quantity of interest which determines Δ_ϵ in (3.38) is the maximum probability of accepting an orthogonal state $|\psi^\perp\rangle$:

$$q := \|\Pi\Omega\Pi\| = \max_{|\psi^\perp\rangle} \langle\psi^\perp|\Omega|\psi^\perp\rangle. \tag{3.49}$$

If a strategy is augmented with an accent or subscript, the parameter q inherits that accent or subscript.

Lemma 17. *Consider an operator* $0 \leq \Omega \leq \mathbb{1}$, $\Omega|\psi\rangle = |\psi\rangle$ *of the form* $\Omega = (1-\alpha)\Omega_1 + \alpha\mathbb{1}$ *for* $0 \leq \alpha \leq 1$. *Then* $q \geq q_1$.

Proof. For arbitrary $|\psi^\perp\rangle$ such that $\langle\psi|\psi^\perp\rangle = 0$, $\langle\psi^\perp|\Omega|\psi^\perp\rangle = (1-\alpha)\langle\psi^\perp|\Omega_1|\psi^\perp\rangle + \alpha$. This is maximised by choosing $|\psi^\perp\rangle$ such that $\langle\psi^\perp|\Omega_1|\psi^\perp\rangle = q_1$, giving $q = (1-\alpha)q_1 + \alpha \geq q_1$. □

We are now in a position to derive the optimal strategy. Note that the special cases where $|\psi\rangle$ is a product state ($\theta = 0$ or $\frac{\pi}{2}$) or a Bell state ($\theta = \frac{\pi}{4}$) are treated separately.

Theorem 18. *Any optimal strategy for verifying a state of the form* $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$ *for* $0 < \theta < \frac{\pi}{2}$, $\theta \neq \frac{\pi}{4}$ *that accepts* $|\psi\rangle$ *with certainty and satisfies the properties of locality, trust and projective measurement, can be expressed as a strategy involving four measurement settings:*

$$\Omega^{opt} = \frac{2-\sin(2\theta)}{4+\sin(2\theta)} P^+_{ZZ} + \frac{2(1+\sin(2\theta))}{3(4+\sin(2\theta))} \sum_{k=1}^{3} (\mathbb{1} - |\phi_k\rangle\langle\phi_k|), \tag{3.50}$$

where the states $|\phi_k\rangle$ *are*

$$|\phi_1\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{3.51}$$

$$|\phi_2\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{3.52}$$

$$|\phi_3\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{1}{\sqrt{1+\cot\theta}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle - \frac{1}{\sqrt{1+\cot\theta}}|1\rangle\right). \tag{3.53}$$

The number of measurements needed to verify to within fidelity ϵ *and statistical power* $1-\delta$ *is*

$$n_{opt} \approx (2 + \sin\theta\cos\theta)\epsilon^{-1}\ln\delta^{-1}. \tag{3.54}$$

Proof. We will break the proof approach down into five steps: (i) elucidation of the most general possible measurement strategy for a state of two qubits; (ii) restriction given that every measurement setting must accept $|\psi\rangle$ with certainty; (iii) restriction given that the strategy must share the same symmetries as $|\psi\rangle$; (iv) restriction given some convexity arguments about the space of allowed strategies; and (v) parameterisation and optimisation over the remaining strategy.

(i) The most general strategy for a state of two qubits

The strategy Ω can be written as a convex combination of local projectors. We can group the projectors by their action according to two local parties, Alice and Bob, and then it must be expressible as a convex combination of five types of terms, grouped by trace:

$$\Omega = c_1 \sum_i \mu_i (\rho_1^i \otimes \sigma_1^i) + c_2 \sum_j \nu_j (\rho_2^j \otimes \sigma_2^j + \rho_2^{j\perp} \otimes \sigma_2^{j\perp}) + c_3 \sum_k \eta_k (\mathbb{1} - \rho_3^k \otimes \sigma_3^k) + c_4 \sum_l [\zeta_l (\rho_4^l \otimes \mathbb{1}) + \xi_l (\mathbb{1} \otimes \sigma_4^l)] + c_5 \mathbb{1} \otimes \mathbb{1}, \tag{3.55}$$

where ρ_i^k and σ_i^k are single-qubit pure states and the subscript denotes the type of term in question. The state $\rho^{j\perp}$ is the density matrix defined by $\mathrm{tr}(\rho^j \rho^{j\perp}) = 0$. Qualitatively, given two local parties Alice and Bob with access to one qubit each, and projectors with outcomes $\{\lambda, \bar{\lambda}\}$, the terms above correspond to the following strategies: (1) Alice and Bob both apply a projective measurement and accept if both outcomes are λ; (2) Alice and Bob both apply a projective measurement and accept if both outcomes agree; (3) Alice and Bob both apply a projective measurement and accept unless both outcomes are λ; (4) Alice or Bob applies a projective measurement and accepts on outcome λ, and the other party abstains; and (5) both Alice and Bob accept without applying a measurement.

(ii) Restriction given that the target state is accepted with certainty

We show in § 2.1.6 that strategies that accept $|\psi\rangle$ with certainty have a quadratic advantage in scaling in terms of ϵ. Given this, we enforce this constraint from the outset and then show that a viable strategy can still be constructed. For the general strategy in Eq. 3.55 to accept $|\psi\rangle$ with certainty, each term in its expansion must accept $|\psi\rangle$ with certainty. However, this is impossible to achieve for some of the terms in the above expansion. In particular, we show that the terms $(\rho \otimes \sigma)$, $(\rho \otimes \mathbb{1})$ and $(\mathbb{1} \otimes \sigma)$ cannot accept $|\psi\rangle$ with certainty, and the form of the term $(\rho \otimes \sigma + \rho^\perp \otimes \sigma^\perp)$ is restricted.

$(\rho \otimes \sigma)$: given that ρ and σ are pure, write $\rho \otimes \sigma = |u\rangle\langle u| \otimes |v\rangle\langle v|$, and so this term only accepts $|\psi\rangle$ with certainty if $\|(|u\rangle\langle u| \otimes |v\rangle\langle v|)|\psi\rangle\| = 1$. However, for $0 < \theta < \frac{\pi}{2}$ the state $|\psi\rangle$ is entangled and this condition cannot be satisfied.

$(\rho \otimes \mathbb{1})$ or $(\mathbb{1} \otimes \sigma)$: For the term $(\rho \otimes \mathbb{1})$, re-express ρ in terms of its Pauli expansion: $\rho \otimes \mathbb{1} = \frac{1}{2}(\mathbb{1} + \alpha X + \beta Y + \gamma Z) \otimes \mathbb{1}$, for $-1 \leq \alpha, \beta, \gamma \leq 1$. Then the condition that this term accepts with probability $p = 1$ is

$$\langle\psi| \frac{1}{2}(\mathbb{1} + \alpha X + \beta Y + \gamma Z) \otimes \mathbb{1} |\psi\rangle = 1. \tag{3.56}$$

By inserting the definition of $|\psi\rangle$, this becomes $\frac{1}{2}(1 - \gamma \cos(2\theta)) = 1$, which is unsatisfiable for $0 < \theta < \frac{\pi}{2}$. It is readily checkable that an identical condition is derived for the term $\mathbb{1} \otimes \sigma$, given the symmetry of the state $|\psi\rangle$ under swapping.

$(\rho\otimes\sigma+\rho^{\perp}\otimes\sigma^{\perp})$: for this term, we can expand both ρ and σ in terms of Pauli operators:

$$\rho=\frac{1}{2}(\mathbb{1}+\alpha X+\beta Y+\gamma Z);\qquad \rho^{\perp}=\frac{1}{2}(\mathbb{1}-\alpha X-\beta Y-\gamma Z) \tag{3.57}$$

$$\sigma=\frac{1}{2}(\mathbb{1}+\alpha' X+\beta' Y+\gamma' Z);\quad \sigma^{\perp}=\frac{1}{2}(\mathbb{1}-\alpha' X-\beta' Y-\gamma' Z). \tag{3.58}$$

Inserting these expressions and the definition of $|\psi\rangle$ into the condition that $p=1$ gives the constraint

$$\gamma\gamma'+(\alpha\alpha'-\beta\beta')\sin(2\theta)=1. \tag{3.59}$$

Now, we know from the Cauchy-Schwarz inequality that

$$\gamma\gamma'+(\alpha\alpha'-\beta\beta')\sin(2\theta)\leq\sqrt{\alpha'^2+\beta'^2+\gamma'^2}\sqrt{\alpha^2\sin^2(2\theta)+\beta^2\sin^2(2\theta)+\gamma^2}\leq 1, \tag{3.60}$$

where the second inequality is derived from the fact that $\{\alpha,\beta,\gamma\}$, $\{\alpha',\beta',\gamma'\}$ are the parameterisation of a pair of density matrices. There are two ways that this inequality can be saturated: (a) $\sin(2\theta)=1$; (b) $\alpha\alpha'-\beta\beta'=0$, $\gamma\gamma'=1$. In all other cases, the inequality is strict. Thus the constraint in Eq. 3.59 cannot be satisfied in general. Exception (a) corresponds to $\theta=\frac{\pi}{4}$, which is omitted from this proof and treated separately. In exception (b), we have that $\gamma\gamma'=1$ and so either $\gamma=\gamma'=1$ or $\gamma=\gamma'=-1$. In both cases we have that

$$\rho\otimes\sigma+\rho^{\perp}\otimes\sigma^{\perp}=\left(\frac{\mathbb{1}+Z}{2}\otimes\frac{\mathbb{1}+Z}{2}\right)+\left(\frac{\mathbb{1}-Z}{2}\otimes\frac{\mathbb{1}-Z}{2}\right)=P^{+}_{ZZ}, \tag{3.61}$$

where P^{+}_{ZZ} is the projector onto the positive eigenspace of ZZ. This is the only possible choice for this particular term that accepts $|\psi\rangle$ with certainty.

We can also make use of Lemma 17 to remove the term $\mathbb{1}\otimes\mathbb{1}$. Given this and the restrictions above from enforcing that $p=1$, the measurement strategy can be written

$$\Omega=\alpha P^{+}_{ZZ}+(1-\alpha)\sum_k\eta_k(\mathbb{1}-\rho_k\otimes\sigma_k), \tag{3.62}$$

where $\sum_k\eta_k=1$ and $0\leq\alpha\leq 1$.

(iii) Restriction given symmetries of the target state

We'll try to further narrow down the form of this strategy by *averaging*; i.e. by noting that, as $|\psi\rangle$ is an eigenstate of a matrix $M_\zeta\otimes M_{-\zeta}$ where

$$M_\zeta=\begin{pmatrix}1 & 0\\ 0 & e^{-i\zeta}\end{pmatrix}, \tag{3.63}$$

then conjugating the strategy by $M_\zeta\otimes M_{-\zeta}$ and integrating over all possible ζ cannot make the strategy worse; if we consider an averaged strategy $\langle\Omega\rangle$ such that

$$\langle\Omega\rangle=\frac{1}{2\pi}\int_{-\pi}^{\pi}d\zeta(M_\zeta\otimes M_{-\zeta})\Omega(M_{-\zeta}\otimes M_\zeta), \tag{3.64}$$

then necessarily the performance of $\langle\Omega\rangle$ cannot be worse than that of Ω. To see this, note that the averaging procedure does not affect the probability of accepting the state $|\psi\rangle$. However, for each particular value of ζ the optimisation for the adversary may necessarily lead to different choices for the orthogonal states $|\psi^\perp(\zeta)\rangle$, and so averaging over ζ cannot be better for the adversary than choosing the optimal $|\psi^\perp\rangle$ at $\zeta = 0$.

We can also consider discrete symmetries of the state $|\psi\rangle$. In particular, $|\psi\rangle$ is invariant under both swapping the two qubits, and complex conjugation (with respect to the standard basis); by the same argument, averaging over these symmetries (i.e. by considering $\Omega' = \frac{1}{2}(\Omega + (\mathrm{SWAP})\Omega(\mathrm{SWAP}^\dagger))$ and $\Omega'' = \frac{1}{2}(\Omega + \Omega^*)$) cannot produce strategies inferior to the original Ω. Therefore we can consider a strategy averaged over these families of symmetries of Ω, without any loss in performance.

This averaging process is useful for three reasons. Firstly, it heavily restricts the number of free parameters in Ω requiring optimisation. Secondly, it allows us to be explicit about the general form of Ω. Thirdly, the averaging procedures are distributive over addition; and so we can make the replacement

$$\begin{aligned}\Omega &= \alpha P^+_{ZZ} + (1-\alpha)\sum_k \eta_k(\mathbb{1} - \rho_k \otimes \sigma_k) \to \langle \alpha P^+_{ZZ} + (1-\alpha)\sum_k \eta_k(\mathbb{1} - \rho_k \otimes \sigma_k)\rangle \\ &= \alpha P^+_{ZZ} + (1-\alpha)\sum_k \eta_k \langle \mathbb{1} - \rho_k \otimes \sigma_k\rangle. \end{aligned} \tag{3.65}$$

Note that a single term $\mathbb{1} - \rho_k \otimes \sigma_k$, may, after averaging, be a convex combination of multiple terms of the form $\mathbb{1} - \rho \otimes \sigma$. To proceed, we will use this averaging procedure to show that it suffices to only include a single, post-averaging term of the form $\langle \mathbb{1} - \rho_k \otimes \sigma_k\rangle$ in the strategy Ω, and that the resulting operator can be explicitly decomposed into exactly three measurement settings.

Consider a general operator Ω, expressed as a 4×4 matrix. First, take the discrete symmetries of $|\psi\rangle$. Averaging over complex conjugation in the standard basis implies that the coefficients of $\langle\Omega\rangle$ are real; and averaging over qubit swapping implies that $\langle\Omega\rangle$ is symmetric with respect to swapping of the two qubits. Denote the operator after averaging these discrete symmetries as $\bar{\Omega}$. Then consider averaging over the continuous symmetry of $|\psi\rangle$:

$$\langle\Omega\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\zeta (M_\zeta \otimes M_{-\zeta})\bar{\Omega}(M_{-\zeta}\otimes M_\zeta) \tag{3.66}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} d\zeta \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\zeta} & 0 & 0 \\ 0 & 0 & e^{-i\zeta} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega_{00} & \omega_{01} & \omega_{01} & \omega_{03} \\ \omega_{01} & \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{01} & \omega_{12} & \omega_{11} & \omega_{13} \\ \omega_{03} & \omega_{13} & \omega_{13} & \omega_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\zeta} & 0 & 0 \\ 0 & 0 & e^{i\zeta} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.67}$$

Thus after averaging using the above symmetries of $|\psi\rangle$, $\langle\Omega\rangle$ can be written in the standard basis as

$$\langle\Omega\rangle = \begin{pmatrix} a & 0 & 0 & b \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ b & 0 & 0 & d \end{pmatrix}, \tag{3.68}$$

for $a, b, c, d \in \mathbb{R}$. Enforcing that the strategy accepts $|\psi\rangle$ with certainty yields $\langle\Omega\rangle|\psi\rangle = |\psi\rangle$, or explicitly that

$$\langle\Omega\rangle = \begin{pmatrix} 1 - b\cot\theta & 0 & 0 & b \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ b & 0 & 0 & 1 - b\tan\theta \end{pmatrix}. \tag{3.69}$$

The eigensystem of this operator is then completely specified; besides $|\psi\rangle$, it has the following eigenvectors:

$$|v_1\rangle = \cos\theta|00\rangle - \sin\theta|11\rangle; \quad |v_2\rangle = |01\rangle; \quad |v_3\rangle = |10\rangle, \tag{3.70}$$

with corresponding eigenvalues $\lambda_1 = 1 - b\csc\theta\sec\theta$ and $\lambda_2 = \lambda_3 = c$. The maximum probability of accepting a state orthogonal to $|\psi\rangle$, q, can then be written

$$q = \|\Pi\langle\Omega\rangle\Pi\| = \max\{\lambda_1, \lambda_2\}, \tag{3.71}$$

where $\Pi = \mathbb{1} - |\psi\rangle\langle\psi|$. Therefore, any reasoning about q can be reduced to reasoning about the pair (λ_1, λ_2).

(iv) Restriction from convexity arguments

We will show that it suffices to only consider a single term of the form $\langle\mathbb{1} - \rho_k \otimes \sigma_k\rangle$ in the decomposition of Ω. We write a strategy of this form as

$$\Omega = \alpha P^+_{ZZ} + (1 - \alpha)\langle\mathbb{1} - \rho \otimes \sigma\rangle. \tag{3.72}$$

For the term $\langle\mathbb{1} - \rho \otimes \sigma\rangle$, we have a constraint on the trace; if we label the eigenvalues for this term as $\lambda_1^{(3)}$ and $\lambda_2^{(3)}$, we have the constraint that $1 + \lambda_1^{(3)} + 2\lambda_2^{(3)} = \mathrm{tr}\langle\mathbb{1} - \rho \otimes \sigma\rangle = 3 \Rightarrow \lambda_2^{(3)} = 1 - \frac{\lambda_1^{(3)}}{2}$. The locus of points satisfying this constraint is plotted in the (λ_1, λ_2) plane as the thick black line in Fig. 3.1. Moreover, we will show that a single term of this form can achieve any valid choice of $\lambda_1^{(3)}$ on this locus (which we defer until we have an explicit parameterisation of terms of this type; see Eq. 3.84, below).

However, we also have an additional constraint derived from insisting that the strategy remains local. For example, the point $(0, 1)$ in the (λ_1, λ_2) plane represents the strategy $\Omega = \mathbb{1} - |v_1\rangle\langle v_1|$, which corresponds to the strategy where the verifier

projects onto $|v_1\rangle$ and accepts if the outcome is not $|v_1\rangle$. But this type of measurement is operationally forbidden as $|v_1\rangle$ is entangled.

It can be readily checked that, for an arbitrary θ, it is not possible to cover the full locus in the range $0 \leq \lambda_1 \leq 1$ with a separable strategy; instead, there is a fixed lower bound on $\lambda_1^{(3)}$. To see this, write

$$\langle \mathbb{1} - \rho \otimes \sigma \rangle = |\psi\rangle\langle\psi| + \lambda_1^{(3)}|v_1\rangle\langle v_1| + \frac{2-\lambda_1^{(3)}}{2}(|v_2\rangle\langle v_2| + |v_3\rangle\langle v_3|). \tag{3.73}$$

Then, taking just the $\langle \rho \otimes \sigma \rangle$ part and expressing as a matrix in the computational basis gives

$$\langle \rho \otimes \sigma \rangle = \begin{pmatrix} (1-\lambda_1^{(3)})\cos^2\theta & 0 & 0 & (\lambda_1^{(3)}-1)\cos\theta\sin\theta \\ 0 & \frac{\lambda_1^{(3)}}{2} & 0 & 0 \\ 0 & 0 & \frac{\lambda_1^{(3)}}{2} & 0 \\ (\lambda_1^{(3)}-1)\cos\theta\sin\theta & 0 & 0 & (1-\lambda_1^{(3)})\sin^2\theta \end{pmatrix}. \tag{3.74}$$

To enforce separability it is necessary and sufficient to check positivity under partial transposition, yielding the constraint $\lambda_1^{(3)} - (1-\lambda_1^{(3)})\sin(2\theta) \geq 0$. Simple rearrangement gives a lower bound that must be satisfied for the strategy to remain separable:

$$\lambda_1^{(3)} \geq \frac{\sin(2\theta)}{1+\sin(2\theta)} := \lambda_{LB}. \tag{3.75}$$

This additional locality constraint rules out any point on the black line to the left of the red point in Fig. 3.1. The term P_{ZZ}^+ has parameters $\lambda_1^{ZZ} = 1$, $\lambda_2^{ZZ} = 0$ and so represents a single point in the (λ_1, λ_2) plane. Thus the parameters (λ_1, λ_2) for the full strategy Ω must be represented by a point in the convex hull of the single point representing the P_{ZZ}^+ term and the locus of points representing the trace 3 part - i.e. in the unshaded region in Fig. 3.1.

We now show that a strategy that includes more trace 3 terms cannot improve on the performance of the strategy above. Write this expanded strategy as

$$\Omega' = \alpha P_{ZZ}^+ + (1-\alpha)\langle \sum_k \eta_k (\mathbb{1} - \rho_k \otimes \sigma_k) \rangle, \tag{3.76}$$

for $\sum_k \eta_k = 1$. Firstly, we note again that the averaging operations (SWAP, conjugation via M_ζ and complex conjugation in the standard basis) are distributive over addition and so we can make the replacement

$$\Omega' = \alpha P_{ZZ}^+ + (1-\alpha)\sum_k \eta_k \langle \mathbb{1} - \rho_k \otimes \sigma_k \rangle. \tag{3.77}$$

Write the composite term $\sum_k \eta_k \langle \mathbb{1} - \rho_k \otimes \sigma_k \rangle := \Omega_{\text{comp}}$, with parameters λ_1^{comp} and λ_2^{comp}. Note that each term in Ω_{comp} satisfies both the constraint from the trace and

the constraint from PPT in Eq. 3.75, and hence so does Ω_{comp}. Now, each operator in this term shares the same eigenbasis (namely, the set of states $\{|v_i\rangle\}$ in Eq. 3.70). Thus we know that $\lambda_1^{\text{comp}} = \sum_k \eta_k \lambda_{1,k}$, and likewise for λ_2^{comp}; i.e. the strategy parameters for this composite term are just a convex combination of those for its constituent parts. A term Ω_{comp} is then specified in the (λ_1, λ_2) plane by a point $\mathcal{P}_{\text{comp}} = (\lambda_1^{\text{comp}}, \lambda_2^{\text{comp}}) \in \text{Conv}(\lambda_{1,k}, \lambda_{2,k})$ (i.e. the point $\mathcal{P}_{\text{comp}}$ must lie on the thick black line bounding the unshaded region in Fig. 3.1).

Thus we know that $\text{Conv}(\Omega') \subseteq \text{Conv}(\Omega)$, and so any strategy writable in the form in Eq. 3.76 can be replaced by a strategy of the form in Eq. 3.72 with identical parameters (λ_1, λ_2), and hence identical performance. Thus, we need only consider strategies of the form

$$\Omega = \alpha P_{ZZ}^{+} + (1-\alpha)\langle \mathbb{1} - \rho \otimes \sigma \rangle. \tag{3.78}$$

(v) Parameterisation and optimisation of the remaining strategy

We can now be explicit about the form of the above strategy. For Ω to accept $|\psi\rangle$ with certainty, $\rho \otimes \sigma$ must annihilate $|\psi\rangle$ and so we make the replacement $\rho \otimes \sigma = |\tau\rangle\langle\tau|$, where $|\tau\rangle$ is the most general pure product state that annihilates $|\psi\rangle$. To be explicit about the form of the state $|\tau\rangle$, write a general two-qubit separable state as

$$|\tau\rangle = (\cos\phi|0\rangle + e^{i\eta}\sin\phi|1\rangle) \otimes (\cos\xi|0\rangle + e^{i\zeta}\sin\xi|1\rangle), \tag{3.79}$$

where we take $0 \leq \phi, \xi \leq \frac{\pi}{2}$, without loss of generality. The constraint that this state annihilates $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$ is

$$\cos\phi\cos\xi\sin\theta + e^{-i(\eta+\zeta)}\sin\phi\sin\xi\cos\theta = 0. \tag{3.80}$$

If either $\phi = 0$ or $\xi = 0$, then $\cos\phi\cos\xi\sin\theta = 0$ implying that $\xi = \frac{\pi}{2}$ or $\phi = \frac{\pi}{2}$, respectively. This yields the annihilating states $|\tau\rangle = |01\rangle$ and $|\tau\rangle = |10\rangle$, respectively. If $\phi, \xi \neq 0$ then from the imaginary part of Eq. 3.80 we find that $e^{-i(\eta+\zeta)} = -1$. Then we can rearrange to give

$$\tan\phi\tan\xi = \tan\theta. \tag{3.81}$$

Using this constraint and the identities

$$\cos\xi = \frac{1}{\sqrt{1+\tan^2\xi}}; \quad \sin\xi = \frac{\tan\xi}{\sqrt{1+\tan^2\xi}}, \tag{3.82}$$

we can eliminate ξ to yield

$$|\tau\rangle = (\cos\phi|0\rangle + e^{i\eta}\sin\phi|1\rangle) \otimes \left(\frac{\tan\phi}{\sqrt{\tan^2\phi + \tan^2\theta}}|0\rangle - \frac{e^{-i\eta}\tan\theta}{\sqrt{\tan^2\phi + \tan^2\theta}}|1\rangle \right). \tag{3.83}$$

Note that, for $0 < \theta < \frac{\pi}{2}$, taking the limits $\phi \to 0$ and $\phi \to \frac{\pi}{2}$ we recover the cases $|\tau\rangle = |01\rangle$ and $|\tau\rangle = |10\rangle$, up to irrelevant global phases. Thus we can proceed without loss of generality by assuming that $\rho \otimes \sigma = |\tau\rangle\langle\tau|$, where $|\tau\rangle$ is given by Eq. 3.83. Averaging over the symmetries of $|\psi\rangle$ outlined above then yields the following expression:

$$\langle \rho \otimes \sigma \rangle = \frac{1}{t^2\phi + t^2\theta} \begin{pmatrix} s^2\phi & 0 & 0 & -s^2\phi t\theta \\ 0 & \frac{1}{2}\left(c^2\phi t^2\theta + s^2\phi t^2\phi\right) & 0 & 0 \\ 0 & 0 & \frac{1}{2}\left(c^2\phi t^2\theta + s^2\phi t^2\phi\right) & 0 \\ -s^2\phi t\theta & 0 & 0 & s^2\phi t^2\theta \end{pmatrix}, \tag{3.84}$$

using the shorthand s, c, t for sin, cos and tan, respectively. Given this explicit parameterisation we can extract the eigenvalue $\lambda_1^{(3)}$:

$$\lambda_1^{(3)} = 1 - \frac{\sec^2\theta \sin^2\phi}{\tan^2\theta + \tan^2\phi}. \tag{3.85}$$

It can be shown by simple differentiation w.r.t. ϕ that, for fixed θ, this expression has a minimum at $\lambda_1^{(3)} = \lambda_{LB}$. Also, this expression is a continuous function of ϕ and therefore can take any value up to its maximum (namely, 1). Hence a single trace 3 term is enough to achieve any point in the allowable convex hull in Fig. 3.1. For convenience we will denote $\tan^2\phi = P$, $\tan^2\theta = T$ for $0 \le P \le \infty$, $0 < T < \infty$. The explicit form for the whole strategy is then

$$\Omega = \begin{pmatrix} \frac{T+P(P+T+\alpha)}{(1+P)(P+T)} & 0 & 0 & \frac{(1-\alpha)P\sqrt{T}}{(1+P)(P+T)} \\ 0 & \frac{(1-\alpha)(T+2P+P^2+2PT)}{2(1+P)(P+T)} & 0 & 0 \\ 0 & 0 & \frac{(1-\alpha)(T+2P+P^2+2PT)}{2(1+P)(P+T)} & 0 \\ \frac{(1-\alpha)P\sqrt{T}}{(1+P)(P+T)} & 0 & 0 & \frac{T+P(1+P+\alpha T)}{(1+P)(P+T)} \end{pmatrix}. \tag{3.86}$$

We now optimise over the two remaining free parameters, $\{\alpha, \phi\}$ (or alternatively, $\{\alpha, P\}$) for fixed θ (or fixed T). This optimisation is rather straightforward from inspection (see Fig. 3.2), and the reader may wish to skip to the answer in Eq. 3.93. However, we include an analytic proof for the sake of completeness. We have shown that it suffices to consider the eigenvalues λ_1 and λ_2, given in this case by the expressions

$$\lambda_1(\alpha, P, T) = 1 - \frac{P(1-\alpha)(1+T)}{(1+P)(P+T)}; \quad \lambda_2(\alpha, P, T) = (1-\alpha)\left[1 - \frac{T+P^2}{2(1+P)(P+T)}\right]. \tag{3.87}$$

The parameter q is given by the maximum of these two eigenvalues. Note that, if $P = 0$, the expression $\lambda_1(\alpha, 0, T) = 1$ which implies that the adversary can pick a state that the verifier always accepts, and hence the strategy fails. Likewise, taking the limit $\lim_{P\to\infty} \lambda_1(\alpha, P, T) = 1$. Thus we must restrict to the range $0 < P < \infty$ to construct a viable strategy for the verifier. The quantity q is minimised for fixed T when the

derivatives with respect to P and α vanish. First, we calculate the derivatives w.r.t. α:

$$\frac{\partial\lambda_1}{\partial\alpha}=\frac{P(1+T)}{(1+P)(P+T)};\quad \frac{\partial\lambda_2}{\partial\alpha}=\frac{-(2P+P^2+T+2PT)}{2(1+P)(P+T)}. \tag{3.88}$$

Given that $P>0$ and $T>0$, we have that for any choice of T, $\partial_\alpha\lambda_1>0$ and $\partial_\alpha\lambda_2<0$. Thus, one of three cases can occur: (a) for a given choice of T and P, the lines given by λ_1 and λ_2 intersect in the range $0\leq\alpha\leq 1$ and hence there is a valid α such that q is minimised when $\lambda_1=\lambda_2$; (b) for a given choice of T and P, $\lambda_1>\lambda_2$ in the range $0\leq\alpha\leq 1$ and hence q is minimised when $\alpha=0$; (c) for a given choice of T and P, $\lambda_1<\lambda_2$ in the range $0\leq\alpha\leq 1$ and hence q is minimised when $\alpha=1$. However, we note that this final case cannot occur; it suffices to check that $\lambda_1(\alpha=1)>\lambda_2(\alpha=1)$, and from the expressions in Eq. 3.87 we have that $\lambda_1(\alpha=1)=1$ and $\lambda_2(\alpha=1)=0$. As a visual aid for the remaining two cases, see Fig. 3.2. In case (a),

$$q=\lambda_1=\lambda_2=\frac{1}{2}+\frac{1}{2}\left(\frac{T+P^2}{T+P^2+4P(1+T)}\right). \tag{3.89}$$

In case (b), we have that

$$q=\lambda_1(0,P,T)=\frac{T+P^2}{(1+P)(P+T)}. \tag{3.90}$$

We must also minimise w.r.t. ϕ; however, we can safely minimise w.r.t. P as $\partial_\phi P>0$ (unless $\phi=0$, but in this case $q=1$ and the strategy fails). In case (b), we have

$$\frac{\partial q}{\partial P}=\frac{(P^2-T)(1+T)}{(1+P)^2(P+T)^2}. \tag{3.91}$$

In this case, consider the two points implicitly defined by the constraint $\lambda_1(0,P,T)=\lambda_2(0,P,T)$ (drawn as the black points in Fig. 3.2). Denote these points $f^{\pm}(T)$. It can be readily checked that in case (b), $\partial_P q<0$ for any $q<f^-(T)$, and $\partial_P q>0$ for any $q>f^+(T)$. Thus the minimum w.r.t P must occur when $\lambda_1(0,P,T)=\lambda_2(0,P,T)$ and hence we can restrict our attention to case (a) (note Fig. 3.2). In this case, $\partial_P q$ becomes

$$\frac{\partial q}{\partial P}=\frac{-2(1+T)(T-P^2)}{[T+4PT+P(4+P)]^2}=0, \tag{3.92}$$

which implies that $P=\sqrt{T}$. Substituting in the optimal choices for the parameters $\{\alpha,P\}$ and re-expressing solely in terms of θ gives the optimal strategy

$$\Omega^{opt}=\frac{2-\sin(2\theta)}{4+\sin(2\theta)}P^+_{ZZ}+\frac{2(1+\sin(2\theta))}{4+\sin(2\theta)}\Omega_3^{opt}, \tag{3.93}$$

where Ω_3^{opt} is given by

$$\Omega_3^{opt}=\mathbb{1}-\frac{1}{(1+t)^2}\begin{pmatrix}1&0&0&-t\\0&t&0&0\\0&0&t&0\\-t&0&0&t^2\end{pmatrix},\quad t=\tan\theta. \tag{3.94}$$

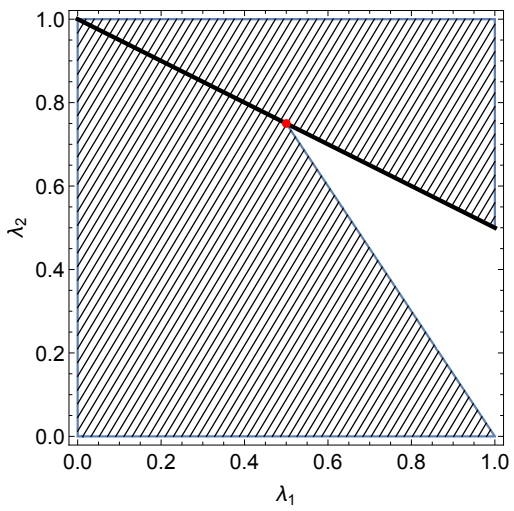

Figure 3.1: Shaded region: unreachable parameters given a strategy Ω that is both local and of the form $\Omega = \alpha P^{+}_{ZZ} + (1-\alpha)\Omega_3$, where Ω_3 is the trace 3 part. Here, $\theta = \frac{\pi}{8}$.

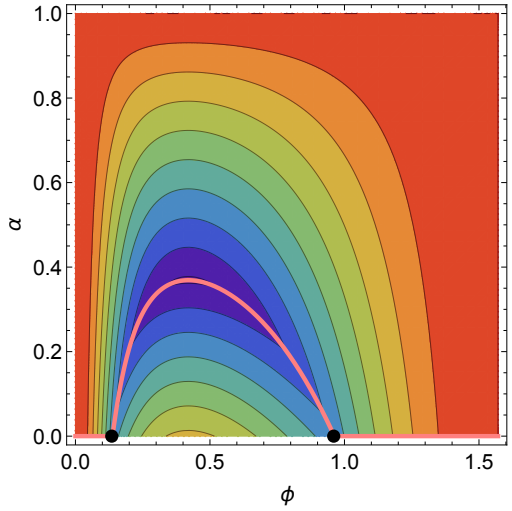

Figure 3.2: A contour map of the function $q(\alpha,\phi) = \max\{\lambda_1(\alpha,\phi), \lambda_2(\alpha,\phi)\}$ for $\theta = \frac{\pi}{8}$, where the pair (λ_1, λ_2) are given in 3.87. The pink curve denotes the minimum w.r.t α given fixed ϕ. Above the curve, $\lambda_1 > \lambda_2$; below, $\lambda_1 < \lambda_2$.

This strategy accepts an orthogonal state with probability

$$q_{opt} = \frac{2+\sin(2\theta)}{4+\sin(2\theta)}, \tag{3.95}$$

implying that the number of measurements needed to verify to within accuracy ϵ and with statistical power $1-\delta$ under this test is

$$n_{opt} = \frac{\ln\delta^{-1}}{\ln((1-\Delta_\epsilon)^{-1})} = \frac{\ln\delta^{-1}}{\ln((1-\epsilon(1-q^{opt}))^{-1})} \approx (2+\sin\theta\cos\theta)\epsilon^{-1}\ln\delta^{-1}. \tag{3.96}$$

The final step is to decompose the operator Ω_3^{opt} into a small set of local, projective measurements. We can do so with a strategy involving only three terms:

$$\Omega_3^{opt} = \frac{1}{3}\left[\sum_{k=1}^{3}(\mathbb{1} - |\phi_k\rangle\langle\phi_k|)\right], \tag{3.97}$$

where the set of separable states $\{|\phi_k\rangle\}$ are the following:

$$|\phi_1\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right)\otimes\left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{3.98}$$

$$|\phi_2\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right)\otimes\left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{3.99}$$

$$|\phi_3\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{1}{\sqrt{1+\cot\theta}}|1\rangle\right)\otimes\left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle - \frac{1}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{3.100}$$

which gives a strategy of the required form. □

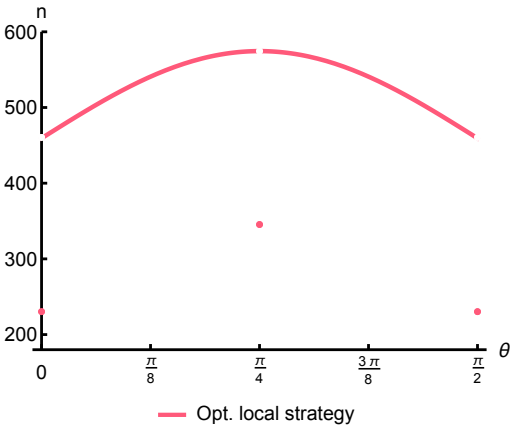

Figure 3.3: The number of measurements needed to verify the state $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, as a function of θ, using the optimal strategy. See Eq. 3.54. Here, $1-\epsilon = 0.99$ and $1-\delta = 0.9$.

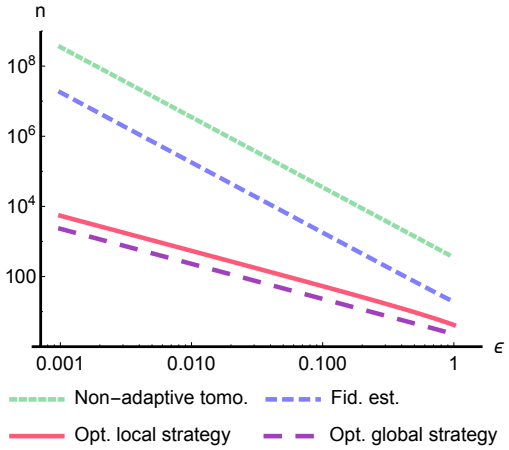

Figure 3.4: A comparison of the total number of measurements required to verify up to fidelity $1-\epsilon$ for the strategy derived here, versus the known bounds for estimation up to fidelity $1-\epsilon$ using non-adaptive tomography in [213] and the fidelity estimation protocol in [84], and the globally optimal strategy given by projecting onto $|\psi\rangle$. Here, $1-\delta = 0.9$ and $\theta = \frac{\pi}{8}$.

We now briefly treat the special cases that were omitted from the above proof: $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}$. In these cases, $|\psi\rangle$ admits a wider choice of measurements that accept with certainty.

$\theta = 0, \theta = \frac{\pi}{2}$: In these cases, the state $|\psi\rangle = |00\rangle$ or $|\psi\rangle = |11\rangle$. Then the globally optimal strategy, just projecting onto $|\psi\rangle$, is an allowed local measurement. Thus in these cases the optimal strategy is to just apply the projector $|00\rangle\langle 00|$ or $|11\rangle\langle 11|$. Given this strategy we have that $p = 1$ and $q = 0$, giving a scaling of the number of measurements required as

$$n_{opt} \approx \epsilon^{-1} \ln \delta^{-1}. \tag{3.101}$$

$\theta = \frac{\pi}{4}$: This case is treated explicitly in § 3.3. The optimal strategy is to perform the Pauli measurements XX, $-YY$ and ZZ with equal weight; i.e.

$$\Omega = \frac{1}{3}(P^+_{XX} + P^+_{-YY} + P^+_{ZZ}), \tag{3.102}$$

where P^+_M is the projector onto the positive eigensubspace of the operator M. In this case, the number of measurements required is

$$n_{opt} \approx \frac{3}{2}\epsilon^{-1} \ln \delta^{-1}. \tag{3.103}$$

We note that this leads to a discontinuity in the number of measurements needed as a function of θ, for fixed ϵ (as seen in Fig. 3.3). This arises since our strategies are designed to have the optimal scaling $\left(O\left(\frac{1}{\epsilon}\right)\right)$ for fixed θ, achieved by having strategies that accept $|\psi\rangle$ with probability 1.

As for scaling, in Fig. 3.4 the number of measurements required to verify a particular two-qubit state of this form, for three protocols, is shown. The optimal protocol derived here gives a marked improvement over the previously published bounds for both tomography [213] and fidelity estimation [84] for the full range of ϵ, for the given values of θ and δ. The asymptotic nature of the advantage for the protocol described here implies that the gap between the optimal scheme and tomography only grows as the requirement on ϵ becomes more stringent. Note also that the optimal local strategy is only marginally worse than the best possible strategy of just projecting onto $|\psi\rangle$.

3.5 Verifying stabilizer states

We now discuss verification strategies for stabilizer states. We take $|\psi\rangle$ to be a stabilizer state of N qubits, namely that there exists a generating set of N commuting Pauli operators $M_1,\ldots,M_N$ on N qubits such that $M_i|\psi\rangle = |\psi\rangle$ for all i. Stabilizer states are ubiquitous in various areas of quantum information, for example in quantum error correction and measurement-based quantum computing; for an introduction to the stabilizer formalism, see [90, 89] and [170], § 10.5. We will describe below a strategy constructed from only stabilizer measurements that accepts $|\psi\rangle$ with certainty, and hence achieves the same asymptotic scaling in the number of required measurements with respect to ϵ as the two-qubit case above. However, we do not rule out that there may be non-stabilizer strategies that give a small constant factor improvement over the strategy defined here.

Theorem 19. *Write a stabilizer state $|\psi\rangle$ and strategy $\Omega = \sum_{j=1}^{K} \mu_j P_j$, where the set $\{P_j\}$ are the projectors onto the positive eigenspace of K linearly independent stabilizers of $|\psi\rangle$, for $K \leq 2^N - 1$. Then the optimal choice of the parameter K and weights μ_j are $K = 2^N - 1$; $\mu_j = \frac{1}{2^N-1}$ for all j. The number of measurements needed to verify to within fidelity ϵ and statistical power $1-\delta$ is then*

$$n_{opt}^{stab} \approx \frac{2^N - 1}{2^{(N-1)}} \epsilon^{-1} \ln \frac{1}{\delta}. \tag{3.104}$$

Proof. Recall that as the verifier accepts $|\psi\rangle$ with certainty, we are concerned with

the optimisation of Δ_ϵ, which can be written as

$$\Delta_\epsilon = \max_{\Omega} \min_{|\psi^\perp\rangle} \epsilon(1 - \langle\psi^\perp|\Omega|\psi^\perp\rangle) \tag{3.105}$$

$$= \epsilon(1 - \min_{\Omega} \max_{|\psi^\perp\rangle} \langle\psi^\perp|\Omega|\psi^\perp\rangle), \tag{3.106}$$

where the maximisation is over positive matrices Ω such that $\Omega|\psi\rangle = |\psi\rangle$.

Now consider Ω written as a matrix in the basis $\{|\psi\rangle, |\psi_j^\perp\rangle\}$, $j = 1\ldots(2^N - 1)$ where the states $|\psi_j^\perp\rangle$ are mutually orthogonal and all orthogonal to $|\psi\rangle$. Given that $\Omega|\psi\rangle = |\psi\rangle$, we know that $\langle\psi_j^\perp|\Omega|\psi\rangle = 0\ \forall j$. Then in this basis Ω can be written

$$\Omega = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{M} \end{pmatrix}, \tag{3.107}$$

where $\mathbf{0}$ is the $(2^N - 1)$-dimensional zero vector and $\mathbf{M}$ is a $(2^N - 1) \times (2^N - 1)$ Hermitian matrix. Then Ω must be writable as $\Omega = |\psi\rangle\langle\psi| + \sum_{j=1}^{2^N-1} \nu_j |\phi_j\rangle\langle\phi_j|$, where $\sum_j \nu_j |\phi_j\rangle\langle\phi_j|$ is the spectral decomposition of $\mathbf{M}$. Given this decomposition, the optimisation for the adversary is straightforward – pick $|\psi^\perp\rangle$ to be the eigenstate in the decomposition of $\mathbf{M}$ with largest eigenvalue: $|\psi^\perp\rangle = |\phi_{max}\rangle$ where $\nu_{max} = \max_j \nu_j$. Then

$$\Delta_\epsilon = \epsilon(1 - \min_{\Omega} \langle\phi_{max}|\Omega|\phi_{max}\rangle) = \epsilon(1 - \min_{\Omega} \nu_{max}). \tag{3.108}$$

Given this choice by the adversary, the verifier is then forced to set the strategy such that all the eigenvalues of $\mathbf{M}$ are equal; i.e. that $\mathbf{M} = a\mathbb{1}$ for some constant a. To see this, consider an alternative strategy where the eigenvalues ν_j are not equal. Now, consider rewriting Ω in terms of stabilizers of $|\psi\rangle$. For any stabilizer (i.e. tensor product of Paulis, perhaps with an overall phase) M over N qubits, the projector onto the positive eigensubspace has $\mathrm{tr}(P_M^+) = 2^{N-1}$. Given that Ω is built from a convex combination of these projectors, and recalling from Lemma 17 that Ω does not contain an identity term, we also know that $\mathrm{tr}(\Omega) = 2^{N-1}$. However, we have also expanded Ω as $\Omega = |\psi\rangle\langle\psi| + \sum_j \nu_j |\phi_j\rangle\langle\phi_j|$, and so

$$\mathrm{tr}(\Omega) = 1 + \sum_j \nu_j = 2^{N-1}. \tag{3.109}$$

Then, it is straightforward to see that decreasing any eigenvalue below a must result in an increase in at least one other eigenvalue in order to maintain this equality, and hence would increase the value of ν_{max}. Thus the optimal choice for the verifier is to set $\Omega = |\psi\rangle\langle\psi| + a\mathbb{1}^\perp$, where $\mathbb{1}^\perp$ is the identity matrix on the subspace orthogonal to $|\psi\rangle$. Taking the trace of this expression gives

$$\mathrm{tr}[|\psi\rangle\langle\psi| + a\mathbb{1}^\perp] = 1 + (2^N - 1)a = 2^{N-1}. \tag{3.110}$$

This can be rearranged for a and then substituted into the expression for Δ_ϵ, which gives

$$\Delta_\epsilon = \frac{2^{(N-1)}}{2^N - 1}\epsilon, \tag{3.111}$$

or that the number of stabilizer measurements required to verify $|\psi\rangle$ is bounded below by

$$n_{opt}^{stab} \approx \frac{2^N - 1}{2^{(N-1)}}\epsilon^{-1}\ln\delta^{-1}. \tag{3.112}$$

The optimal $\Omega = |\psi\rangle\langle\psi| + \frac{2^{(N-1)}-1}{2^N-1}\mathbb{1}^\perp$ and the optimal scaling above can be achieved by decomposing Ω into a strategy involving a maximal set (excluding the identity) of $2^N - 1$ linearly independent stabilizers, all with equal weight. To see this note that for a stabilizer group of a state $|\psi\rangle$ of N qubits, there are 2^N linearly independent stabilizers (including the identity element). Denote these stabilizers $\{M_i, i = 1 \ldots 2^N\}$. Then, we make use of the fact that [110]

$$\frac{1}{2^N}\sum_{i=1}^{2^N} M_i = |\psi\rangle\langle\psi|. \tag{3.113}$$

Explicitly extracting the identity element gives

$$\sum_{i=1}^{2^N-1} M_i = 2^N|\psi\rangle\langle\psi| - \mathbb{1}. \tag{3.114}$$

Now, each stabilizer (for any N) is a two outcome measurement and so we can make use of the fact that M_i can be written in terms of the projector onto the positive eigenspace of M_i, denoted P_i^+, as $M_i = 2P_i^+ - \mathbb{1}$. Substituting in this expression and rearranging gives

$$\sum_{i=1}^{2^N-1} P_i^+ = 2^{(N-1)}|\psi\rangle\langle\psi| + (2^{(N-1)} - 1)\mathbb{1}. \tag{3.115}$$

Then normalising this expression over $2^N - 1$ stabilizers yields

$$\begin{aligned}\frac{1}{2^N-1}\sum_{i=1}^{2^N-1} P_i^+ &= \frac{2^{(N-1)}}{2^N-1}|\psi\rangle\langle\psi| + \frac{2^{(N-1)}-1}{2^N-1}\mathbb{1} \\ &= \frac{2^{(N-1)} + 2^{(N-1)} - 1}{2^N-1}|\psi\rangle\langle\psi| + \frac{2^{(N-1)}-1}{2^N-1}\mathbb{1}^\perp \\ &= |\psi\rangle\langle\psi| + \frac{2^{(N-1)}-1}{2^N-1}\mathbb{1}^\perp = \Omega,\end{aligned} \tag{3.116}$$

where $\mathbb{1}^\perp$ is the identity matrix on the subspace orthogonal to $|\psi\rangle$, as required. □

Note that for growing N, the quantity n_{opt}^{stab} given in Eq. 3.112 is bounded above by $2\epsilon^{-1}\ln\delta^{-1}$, which does not depend on N, and implies that this stabilizer strategy requires at most a factor of two more measurements than the optimal non-local verification strategy (just projecting onto $|\psi\rangle$).

One could also consider a reduced strategy that involves measuring fewer stabilizers. However, given a state of N qubits and a set of k stabilizers, the dimension of the subspace stabilized by this set is at least 2^{N-k}. Thus for any choice of $k < N$, there must always exist at least one state $|\psi^{\perp}\rangle$ orthogonal to $|\psi\rangle$ that is stabilized by every stabilizer in the set. Then, the adversary can construct a σ that always accepts, implying that the verifier has no discriminatory power between $|\psi\rangle$ and σ and thus the strategy fails. Consider instead constructing a strategy from the N stabilizer generators of $|\psi\rangle$, with corresponding projectors $\{P_j^{s.g.}\}$. Then, $\Omega = \sum_j \mu_j P_j^{s.g.}$. The set of projectors $\{P_j^{s.g.}\}$ commute and so share a common eigenbasis, denoted $\{|\lambda_j\rangle\}$. To optimise this strategy over the weights μ_j, we first need the following lemma:

Lemma 20. *Write the unique sets of $N-1$ independent stabilizer generators of $|\psi\rangle$, $S_k = \{M_j, j = 1\ldots N\} \setminus M_k$, $k = 1\ldots N$. Then each S_k corresponds to a state $|\lambda_k\rangle$, $\langle\lambda_k|\psi\rangle = 0$, such that $\langle\lambda_k|\lambda_l\rangle = \delta_{kl}$.*

Proof. Each set S_k stabilizes a space of dimension two, and so a $|\lambda_k\rangle$ where $\langle\lambda_k|\psi\rangle = 0$ exists. Moreover, the stabilizer generators define an orthogonal eigenbasis of which $|\lambda_k\rangle$ is an element. To show that two sets S_k and S_l, $k \neq l$, define distinct eigenvectors, assume the converse; that $|\lambda_k\rangle \propto |\lambda_l\rangle$. However, then the set $S = S_k \cup S_l$ would stabilize $|\lambda_k\rangle$, which is a contradiction as S is the full set of stabilizer generators and uniquely stabilizes $|\psi\rangle$. □

We can now derive the optimal stabilizer generator strategy.

Theorem 21. *For a stabilizer state $|\psi\rangle$ and strategy $\Omega = \sum_{j=1}^{N} \mu_j P_j^{s.g.}$, where the set $\{P_j^{s.g.}\}$ are the projectors onto the positive eigenspace of the stabilizer generators of $|\psi\rangle$, the optimal choice of the weights μ_j is $\mu_j = \frac{1}{N}$, for all j. The number of measurements needed to verify to within fidelity ϵ and statistical power $1-\delta$ is then*

$$n_{opt}^{s.g.} \approx \frac{N}{\epsilon} \ln \frac{1}{\delta}. \tag{3.117}$$

Proof. If we write a state orthogonal to $|\psi\rangle$ in the stabilizer eigenbasis as $|\psi^{\perp}\rangle = \sum_k \alpha_k |\lambda_k\rangle$, we have that

$$\begin{aligned}
\langle\psi^{\perp}|\Omega|\psi^{\perp}\rangle &= \sum_{k,m=1}^{2^N} \sum_{j=1}^{N} \bar{\alpha}_k \alpha_m \mu_j \langle\lambda_k|P_j^{s.g.}|\lambda_m\rangle \\
&= \sum_{k,m=1}^{2^N} \sum_{j=1}^{N} \bar{\alpha}_k \alpha_m \mu_j \delta_{km} \epsilon_{jk} \\
&= \sum_{k=1}^{2^N} \sum_{j=1}^{N} |\alpha_k|^2 \mu_j \epsilon_{jk} := \sum_{k=1}^{2^N} |\alpha_k|^2 E_k,
\end{aligned} \tag{3.118}$$

where $\epsilon_{jk} = 1$ if $P_j|\lambda_k\rangle = |\lambda_k\rangle$ and zero otherwise. This quantity is the *parity-check matrix* for the set of stabilizers $\{P_j^{s.g.}\}$. The quantity of interest with respect to verification is

$$q = \min_{\Omega} \max_{|\psi^\perp\rangle} \langle\psi^\perp|\Omega|\psi^\perp\rangle = \min_{\mu_j} \max_{\alpha_k} \sum_{j,k} |\alpha_k|^2 \mu_j \epsilon_{jk}, \tag{3.119}$$

where the verifier's minimisation is over the probabilities μ_j with which a stabilizer generator indexed by j is drawn in the protocol, and the adversary maximises over the set of amplitudes α_k that describes the state most likely to fool the verifier. Lemma 20 gives that, from the full set of 2^N basis states $|\lambda_k\rangle$, there is a subset of N basis states $|\lambda_{\tilde{k}}\rangle$, $\tilde{k} \in I$ for $|I| = N$, stabilized by exactly $N-1$ generators; thus for basis states in this subset, the quantity $\epsilon_{j\tilde{k}} = 1 - \delta_{j\tilde{k}}$. Then we can compute the summation over j as

$$E_{\tilde{k}} = \sum_j \mu_j \epsilon_{j\tilde{k}} = \sum_j \mu_j (1 - \delta_{j\tilde{k}}) = 1 - \mu_{\tilde{k}}, \tag{3.120}$$

using the fact that $\sum_j \mu_j = 1$. Now, each element of E_k for $k \notin I$ is a summation of at most $N-2$ terms, μ_j. Thus there always exists another element $E_{\tilde{k}}$ for $\tilde{k} \in I$ that is at least as large; and so it is never detrimental to the adversary to shift any amplitude on the basis state labelled by k to the basis state labelled by $\tilde{k}$. Thus the optimal choice for the adversary's state is $|\psi^\perp\rangle \in \text{span}\{|\lambda_{\tilde{k}}\rangle : \tilde{k} \in I\}$. Given this choice by the adversary, we have that

$$q = \min_{\mu_{\tilde{k}}} \max_{\alpha_{\tilde{k}}} \sum_{\tilde{k}} |\alpha_{\tilde{k}}|^2 (1 - \mu_{\tilde{k}}) = \min_{\mu_{\tilde{k}}} \max_{\tilde{k}} (1 - \mu_{\tilde{k}}). \tag{3.121}$$

It is straightforward to see that the optimal choice for the verifier is to have $\mu_{\tilde{k}} = \frac{1}{N}$, for all $\tilde{k}$; then $\Omega = \frac{1}{N} \sum P_j^{s.g.}$. Thus

$$q = 1 - \frac{1}{N} \Rightarrow n_{opt}^{s.g.} \approx \frac{N}{\epsilon} \ln \frac{1}{\delta}. \tag{3.122}$$

□

In this case, the number of required measurements is $n_{opt}^{s.g.} \approx N\epsilon^{-1} \ln \delta^{-1}$, with this bound saturated by measuring all stabilizer generators with equal weight. Conversely, constructing a measurement strategy from the full set of $2^N - 1$ linearly independent stabilizers requires a number of measurements $n_{opt}^{stab} \approx \frac{2^N - 1}{2^{(N-1)}} \epsilon^{-1} \ln \delta^{-1}$, again with this bound saturated by measuring each stabilizer with equal weight. So clearly, this scaling is much poorer in N than in the case where the full set of $2^N - 1$ linearly independent stabilizers are allowed. For growing N, the latter expression for the number of measurements is bounded from above by $2\epsilon^{-1} \ln \delta^{-1}$, which implies that there is a local strategy for any stabilizer state, of an arbitrary number of qubits, which requires at most twice as many measurements as the optimal

non-local strategy. Note that this strategy may not be exactly optimal; for example, the state $|00\rangle$ is also a stabilizer state, and in this case applying the measurement $|00\rangle\langle 00|$ is both locally implementable and provably optimal. Thus, the exactly optimal strategy may depend more precisely on the structure of the individual state itself. However, the stabilizer strategy is only inferior by a small constant factor. In comparison to the latter strategy constructed from every stabilizer, the former strategy constructed from only the N stabilizer generators of $|\psi\rangle$ has scaling that grows linearly with N. Thus there is ultimately a trade-off between number of measurement settings and total number of measurements required to verify within a fixed fidelity.

3.6 "Soft" verification

One reasonable complaint about the verification game outlined above is that it does not take into account systematic error; i.e. even if the state produced is not far from the target, it may be the case that there is some small systematic error that rules out the ability of the verifier to verify to within an arbitrarily small fidelity. In particular, one could generalise to a "soft" verification scenario where the state is promised to be either (a) within some small ball around the target; or (b) outside some larger ball around the target, and the verifier must decide which. If we take a target state $|\psi\rangle$, experimental outputs ρ_i, and adversarial states σ_i, then the aim is to distinguish between the following scenarios:

1. **(Good)** For the i^{th} run the source outputs ρ_i, for $\langle\psi|\rho_i|\psi\rangle = 1-\epsilon_i' \geq 1-\epsilon'$;

2. **(Bad)** For the i^{th} run the source outputs σ_i, for $\langle\psi|\sigma_i|\psi\rangle = 1-\epsilon_i \leq 1-\epsilon$.

Here ϵ' is a measure of the systematic error; if $\epsilon' = 0$ then the protocol is as previously defined. Clearly, if $\epsilon' \geq \epsilon$ then the protocol must fail, regardless of the choice of strategy by the verifier - the adversary can then pick $\sigma_i = \arg\max_{\rho_j} \epsilon_j'$, $\forall i$, rendering the two hypotheses indistinguishable. We seek to explore the other case, i.e. when $\epsilon' < \epsilon$; in particular, how close the quantities can get before the protocol fails.

In both the two-qubit case in § 3.2.4 and the stabilizer case in § 3.5, the strategy Ω had the property that $\Omega = (1-q)|\psi\rangle\langle\psi| + q\mathbb{1}$, for a parameter q corresponding to the probability of a verifier being fooled by a state orthogonal to $|\psi\rangle$. Then the probability of passing trial i in the "good" case is

$$\mathrm{tr}(\Omega\rho_i) = (1-q)\,\mathrm{tr}(|\psi\rangle\langle\psi|\rho_i) + q = 1-\epsilon_i'(1-q), \tag{3.123}$$

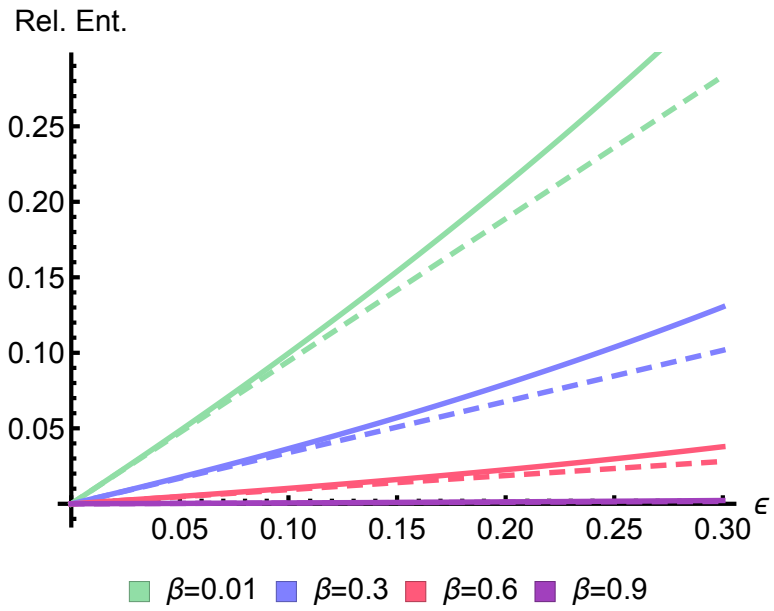

Figure 3.5: A comparison of the relative entropy in the "soft" verification scenario (solid lines) with the first-order approximation (dashed lines). The parameter β is the ratio between the sizes of the infidelity "ball" around the target state in the "good" and "bad" instances. The approximation aligns closely with the true value for a reasonable range of ε.

and likewise the probability of passing in the "bad" case is $1-\epsilon_i(1-q)$. Then the number of copies required to distinguish these two cases is governed by the relative entropy, which in the worst case is bounded by

$$D(1-\epsilon_i'(1-q)\|1-\epsilon_i(1-q)) \geq D(1-\epsilon'(1-q)\|1-\epsilon(1-q)). \tag{3.124}$$

For readability, we'll absorb the constant factor of $1-q$ and define $\bar{\epsilon} := \epsilon(1-q)$, and likewise for ϵ'. Also, we are interested in what happens when ϵ' gets close to ϵ, so let $\beta = \epsilon'/\epsilon$, for $0 \leq \beta < 1$. Then the worst-case relative entropy can be rewritten

$$D(1-\beta\bar{\epsilon}\|1-\bar{\epsilon}) = (1-\beta\bar{\epsilon})\ln\frac{1-\beta\bar{\epsilon}}{1-\bar{\epsilon}} + \beta\bar{\epsilon}\ln\beta. \tag{3.125}$$

Assuming an experiment designed to produce high-fidelity states, we can Taylor expand this expression around $\bar{\epsilon} = 0$ to give

$$D(1-\beta\bar{\epsilon}\|1-\bar{\epsilon}) \approx (1-\beta+\beta\ln\beta)\bar{\epsilon} + O(\bar{\epsilon}^2). \tag{3.126}$$

Taking this series to first order in $\bar{\epsilon}$ gives a very close approximation to the relative entropy, across the entire range of β; see Fig 3.5. Additionally, we can see from Fig. 3.5 that the first-order approximation gives a relatively tight lower bound for the

relative entropy, and hence yields an upper bound on the required number of trials. Thus the number of copies needed to verify in this "soft" scenario is

$$n = \frac{1}{D(1-\beta\bar{\epsilon}\|1-\bar{\epsilon})}\ln\frac{1}{\delta} \leq \frac{1}{1-\beta+\beta\ln\beta}\cdot\frac{1}{(1-q)\epsilon}\ln\frac{1}{\delta}. \tag{3.127}$$

Thus the only difference in complexity between the "soft" scenario and that previously discussed is the prefactor $(1-\beta+\beta\ln\beta)^{-1}$. In the limit where $\beta \to 0$, the ball around the target state reduces to a point, this prefactor tends to 1, and we recover the scaling for the previous scenario as expected. In the limit where $\beta \to 1$, the ball around the target state in the "good" case approaches the ball around the target state in the "bad" case, and the prefactor blows up.

3.7 Verification of arbitrary pure states

The goal of this section is to outline approaches to extending the protocol in § 3.2.4 and § 3.5 to arbitrary pure target states of N qubits.

In contrast to the two-qubit case, there are now three parameters defining the scaling of the number of copies needed to verify a particular pure state $|\psi\rangle \in \mathbb{C}^d$: the confidence δ, the infidelity ϵ and the number of qubits of the target, $N = \log_2(d)$. Typical verification procedures require a number of copies scaling like

$$n = \frac{f(N)}{\epsilon^k}\log\frac{1}{\delta}, \tag{3.128}$$

for some non-decreasing function $f(N)$ and integer k. See Table 2.3 for examples using different verification procedures, including various flavours of tomography. For N fixed, we know that we can achieve a quadratic improvement in ϵ with strategies where every measurement accepts $|\psi\rangle$ with certainty, over those that do not have this property. However, for N growing this may not be the optimal strategy; for example, we might prefer a strategy that scales like $\frac{1}{\epsilon^2}$ if the alternative scales like $\frac{2^N}{\epsilon}$ (for large enough N).

The approaches we take are outlined as follows: in § 3.7.1, we derive projector rank constraints on strategies that accept $|\psi\rangle$ with certainty; in § 3.7.2 we derive a lower bound on the copy complexity when the strategy is constructed from rank 1 projectors; in § 3.7.3 we discuss the shortcomings of directly extending the approach that we took for arbitrary two qubit states; and in § 3.7.4 we outline an ansatz strategy for arbitrary pure states.

3.7.1 Strategy constraints when the target state is accepted with certainty

The first question one might ask is: can we achieve the same scaling with N and ϵ as in the globally optimal strategy (projecting onto $|\psi\rangle$), $n = \frac{1}{\epsilon}\log\frac{1}{\delta}$, with local

measurements? We know that to achieve this scaling with ϵ, we must constrain the strategy so that it always accepts $|\psi\rangle$ with certainty. Given this constraint, we can write Ω in its eigenbasis as

$$\Omega = |\psi\rangle\langle\psi| + \sum_{j=1}^{2^N-1} \nu_j |\psi_j^\perp\rangle\langle\psi_j^\perp|. \tag{3.129}$$

Previously, we have sought out strategies where $\nu_j = q$, for all j, as this is always favourable to the verifier given access to a strategy of fixed rank. q is then the probability that a verifier accepts a state orthogonal to $|\psi\rangle$. Then

$$\Omega = |\psi\rangle\langle\psi| + q \sum_{j=1}^{2^N-1} |\psi_j^\perp\rangle\langle\psi_j^\perp|. \tag{3.130}$$

It is straightforward to show that the goal of a verifier under these constraints is to seek out a valid strategy composed of projectors of smallest possible rank.

Observation 22. *If $\Omega = \sum_j \mu_j P_j$ is of the form in Eq. 3.130, then for $\max_j \mathrm{tr}(P_j) = T$, we have that the number of copies is bounded by*

$$n \lesssim \frac{2^N - 1}{2^N - T} \cdot \frac{1}{\epsilon} \log \frac{1}{\delta}. \tag{3.131}$$

Proof. The strategy is a convex combination of projectors $\Omega = \sum_k \mu_k P_k$ with trace $\mathrm{tr}(P_k) \leq T$, so $\mathrm{tr}\,\Omega \leq T$. Then taking the trace of Eq. 3.130 gives a relation between q and T:

$$\mathrm{tr}(\Omega) = 1 + q(2^N - 1) \leq T \Rightarrow q \leq \frac{T-1}{2^N - 1}. \tag{3.132}$$

Then, the number of copies needed to verify is

$$n = \frac{1}{1-q} \cdot \frac{1}{\epsilon} \log \frac{1}{\delta} \leq \frac{2^N - 1}{2^N - T} \cdot \frac{1}{\epsilon} \log \frac{1}{\delta}. \tag{3.133}$$

□

Thus the scaling with N in this simplified case is dictated by the trace of the projectors in the strategy only (with lower trace yielding a better scaling).

Example 1: if T is a constant independent of N, then

$$\frac{2^N - 1}{2^N - T} = 1 + \frac{T-1}{2^N - T}, \tag{3.134}$$

which asymptotically approaches the scaling for the globally optimal strategy (exponentially quickly).

Example 2: if $T = 2^{N-1}$ (e.g. in the case of stabilizer measurements), then

$$\frac{2^N - 1}{2^N - T} = \frac{2^N - 1}{2^N - 2^{N-1}} = \frac{2^N - 1}{2^{N-1}} = 2 - \frac{1}{2^{N-1}}, \tag{3.135}$$

which is precisely the result previously derived for stabilizer measurements in § 3.5. **Example 3:** if $T = 2^N - C$ for some constant C (e.g. for a strategy constructed from projectors of the form $P_k = \mathbb{1} - |\eta_k\rangle\langle\eta_k|$), then

$$\frac{2^N - 1}{2^N - T} = \frac{2^N - 1}{C}, \tag{3.136}$$

which, while offering no marked improvement over tomography and direct fidelity estimation in terms of scaling with N, still improves the dependence on ϵ quadratically. However, Observation 22 is predicated on Eq. 3.130 being satisfied. It may be the case that a verifier can relax the assumption that $v_j = q$ for all j and then have enough freedom to construct a strategy that scales favourably even with high-rank projectors.

3.7.2 A strategy with restricted-rank measurements

Motivated by Observation 22, one may consider constructing a strategy from the lowest rank (i.e. rank 1) projectors: $\Omega = \sum_j \mu_j |\eta_j\rangle\langle\eta_j|$ for some set of product states $|\eta_j\rangle$. However, in general the target state $|\psi\rangle$ will be entangled, we cannot construct a strategy where every setting accepts the target state with certainty, and we must forfeit the quadratic advantage in ϵ. However, we may wish to proceed, hoping for an advantage in N. In this instance, the strategy Ω is a convex combination of pure states; i.e. it is a density matrix. Consider an Ω of the form

$$\Omega = \frac{1-\Delta}{2^N}\mathbb{1} + \Delta|\psi\rangle\langle\psi|, \tag{3.137}$$

for some N-qubit target state $|\psi\rangle$. The probability that this strategy accepts the target $|\psi\rangle$ is $p = \frac{1-\Delta}{2^N} + \Delta$, and the probability that it accepts a state orthogonal to $|\psi\rangle$ is $q = \frac{1-\Delta}{2^N}$; so the probability that the verifier accepts a state ϵ away from the target is $(1-\epsilon)p + \epsilon q = p - \Delta\epsilon$. The requirement that this strategy be decomposable into locally-implementable projectors is equivalent to the requirement that the density matrix Ω be separable. Luckily, for any state $|\psi\rangle\langle\psi|$ it is already known that there exists a range of sufficiently small Δ, such that any choice in this range yields a separable Ω [32]:

Theorem 23. *For a density matrix of the form*

$$\rho_\Delta = \frac{1-\Delta}{2^N}\mathbb{1} + \Delta\rho, \tag{3.138}$$

ρ_Δ is guaranteed to be separable if

$$0 \leq \Delta \leq \frac{1}{1 + 2^{2N-1}}. \tag{3.139}$$

Assuming that N is sufficiently large, and that Δ is sufficiently small, we can take the Taylor expansion of $D(p\|p-\Delta\epsilon)$ in Eq. 3.23. Then

$$D(p\|p-\Delta\epsilon) \approx \frac{2p(1-p)}{\Delta^2\epsilon^2} = \frac{2(d-1)(1-\Delta)[1+\Delta(d-1)]}{d^2\Delta^2\epsilon^2}, \tag{3.140}$$

for $d = 2^N$. Substituting in the bound for Δ gives

$$D(p\|p-\Delta\epsilon) \leq \frac{2\epsilon^2}{d(d-1)(d+2)}, \tag{3.141}$$

and therefore that the number of copies required is bounded by

$$n \geq \frac{d(d-1)(d+2)}{2\epsilon^2}\log\frac{1}{\delta}. \tag{3.142}$$

While Theorem 23 is not constructive, it is instructive to compare to the known bounds in Table 2.2. The bound on the required number of copies is marginally better in d than in conventional tomography, when there is no a priori guarantee that the target state is pure (but marginally worse if this guarantee exists). On the other hand, this bound offers no advantage over direct fidelity estimation [85].

3.7.3 Extending the two-qubit approach

An obvious avenue to consider when trying to generalise the previous protocol for arbitrary states of two qubits is to adapt the same proof strategy: (i) enumerate all possible types of local measurements that could be implemented, as in Eq. 3.55; (ii) rule out as many as possible that cannot be implemented given the strategy requirements; (iii) parameterise and optimise over the rest.

We can be more quantitative about step (i). For a state of N qubits, a local measurement (in the sense we have discussed above) physically corresponds to N parties, each with a single qubit, able to carry out a two-outcome projective measurement. Thus the total number of possible different measurement patterns is 2^N. Given the set of possible measurement outcome patterns, the parties must choose a subset of measurement patterns on which to accept the state they have (and hence, a complement of patterns on which to reject). Each designation from the set of measurement patterns to "accept" or "reject" is a Boolean function $f : \{0,1\}^N \to \{0,1\}$, and so the number of possible types of measurement is upper bounded by the number of possible Boolean functions over N variables, or 2^{2^N}. However, some of these patterns are physically equivalent. Specifically, each party has the freedom to "flip" their measurement outcomes (i.e. physically to apply the complementary projector $\mathbb{1} - P$, rather than P itself, or mathematically to flip the input bits to the function f), without affecting the protocol. Thus the number of types of measurement of rank T for a state of N qubits, $\mathcal{B}(T,N)$, is the number of

N \ T	0	1	2	3	4	5	6	7	8	⋯	Total
1	1	1	1	-	-	-	-	-	-	-	3
2	1	1	3	1	1	-	-	-	-	-	7
3	1	1	7	7	14	7	7	1	1	-	46
4	1	1	15	35	140	273	553	715	870	⋯	4336
5	1	1	31	155	1240	6293	28861	105183	330640	⋯	134281216

Table 3.1: The number of physically distinct measurement types of rank T applicable to a state of N qubits.

inequivalent Boolean functions with T nonzero values, where the equivalence relation is with respect to action under the "complementary group" (see [103]). Explicitly,

$$\mathcal{B}(T,N) = \begin{cases} 2^{-N}\binom{2^N}{T} & \text{if } T \text{ is odd;} \\ 2^{-N}\left[\binom{2^N}{T} + (2^N-1)\binom{2^{N-1}}{\frac{T}{2}}\right] & \text{if } T \text{ is even.} \end{cases} \tag{3.143}$$

The total number of measurement types for a state of N qubits is

$$\sum_{T=0}^{2^N} \mathcal{B}(T,N) = \frac{2^{2^N} + (2^N-1)2^{2^{N-1}}}{2^N}. \tag{3.144}$$

Numerical values of these expressions for small T and N are shown in Table 3.1. The upshot is that, while feasible for states of two qubits, it rapidly becomes impossible (both computationally and numerically) to proceed with the prior proof strategy for states of more qubits, because the ability to even write down the list of possible measurement types is computationally intractable. Any successful approach must impart sufficient prior knowledge about symmetries in the target state from the outset, in order to strongly reduce the number of possible measurement types to consider.

3.7.4 An arbitrary state ansatz

Suppose instead that we'd like to maintain the quadratic advantage in ϵ for the previously derived strategies in § 3.2.4 and § 3.5, by constructing a strategy from measurements that always except the target (and we'll aim to optimise the scaling of the number of copies with N later). Rather than the axiomatic approach above, we'll take the following ansatz: to verify $|\psi\rangle$, use measurements $\{\mathbb{1} - |\eta_i\rangle\langle\eta_i|, |\eta_i\rangle\langle\eta_i|\}$ where $\langle\eta_i|\psi\rangle = 0$, for all i, to ensure that $p = 1$. We need to explicitly construct a set

$|\eta_i\rangle$ of product states that spans the space orthogonal to $|\psi\rangle$, in order for the strategy to succeed with some non-zero probability. Try the following set of $N \cdot 2^{N-1}$ states:

$$|\phi_{jk}\rangle = \mathrm{SWAP}_{0k}\{(\alpha^*_{jk}|0\rangle + \beta^*_{jk}|1\rangle) \otimes |j\rangle\}, \tag{3.145}$$

where the state $|j\rangle$ runs over all 2^{N-1} computational basis states of $N-1$ qubits, $k = 0 \ldots N-1$, and the sets of parameters α_{jk} and β_{jk} are chosen such that $\langle\phi_{jk}|\psi\rangle = 0$, for all j and k. This constraint for each $|\phi_{jk}\rangle$ is guaranteed to have a solution when $|\psi\rangle$ has support on all computational basis states. If $|\psi\rangle$ does not have support on all computational basis states, then by a result by Montanaro and Shepherd [165] there is always a subset of qubits which, upon operating with Hadamards, results in a state which does have support on all computational basis states. Thus we can always proceed provided that we also remember to conjugate the resultant local strategy by the same set of Hadamards. We now show that the above ansatz forms a valid verification strategy for $|\psi\rangle$.

Theorem 24. *Consider the set of $N \cdot 2^{N-1}$ states on N qubits:*

$$|\phi_{jk}\rangle = SWAP_{0k}\{(\alpha^*_{jk}|0\rangle + \beta^*_{jk}|1\rangle) \otimes |j\rangle\}, \tag{3.146}$$

where j denotes the j^{th} computational basis state over $N-1$ qubits, $j = 0 \ldots 2^{N-1}-1$, $\alpha_{jk}, \beta_{jk} \in \mathbb{C}$, and $k = 1 \ldots N$. If the parameters $\{\alpha_{jk}, \beta_{jk}\}$ are chosen such that $\langle\phi_{jk}|\psi\rangle = 0$ for a state $|\psi\rangle$ that has non-zero amplitude on all computational basis states, then the states $|\phi_{jk}\rangle$ must also span the space orthogonal to $|\psi\rangle$.

Proof. Enforcing that each state $|\phi_{jk}\rangle$ is orthogonal to $|\psi\rangle$ yields $N \cdot 2^{N-1}$ constraints of the form

$$\langle\phi_{jk}|\psi\rangle = \alpha_{jk}\psi_{x_{jk}} + \beta_{jk}\psi_{\bar{x}_{jk}} = 0, \tag{3.147}$$

where the string x_{jk} is bit string j with a 0 inserted at position k, and $\bar{x}_{jk}$ is bit string j with a 1 inserted at position k. Thus these strings differ only at position k (i.e. have Hamming distance 1). If we construct a graph G where each vertex is labelled with $\psi_{x_{jk}}$ and an edge is added between vertices if there exists a constraint of the above form relating their labels, then the resultant graph is an N-dimensional hypercube.

We seek to show that only a single state can satisfy all the constraints in Eq. 3.147. Alternatively, we can consider a matrix M of dimension $(N \cdot 2^{N-1}) \times (2^N)$, where the rows label the constraints in Eq. 3.147 and the columns label amplitudes of the state satisfying the constraints; then

$$M_{(jk),l} = \begin{cases} -\psi_{\bar{x}_{jk}} & \text{if } \; x_{jk} = l \\ \psi_{x_{jk}} & \text{if } \; \bar{x}_{jk} = l \\ 0 & \text{otherwise.} \end{cases} \tag{3.148}$$

Any state $|v\rangle$ whose elements satisfy all the constraints in Eq. 3.147 must equivalently satisfy $M|v\rangle = 0$. Suppose that this is the case; then by checking each row of M, $|v\rangle$ must satisfy

$$-\psi_{\bar{x}_{jk}} v_{x_{jk}} + \psi_{x_{jk}} v_{\bar{x}_{jk}} = 0. \tag{3.149}$$

Alternatively, given that all the amplitudes of $|\psi\rangle$ are non-zero,

$$\frac{v_{x_{jk}}}{\psi_{x_{jk}}} = \frac{v_{\bar{x}_{jk}}}{\psi_{\bar{x}_{jk}}}. \tag{3.150}$$

From the graph G, the ratio of two amplitudes of $|v\rangle$ is equal to the corresponding ratio of amplitudes of $|\psi\rangle$, provided that their vertices are adjacent. However, we may transition along a path in G consisting of multiple edges and concatenate ratios; and given that G is connected, we can reach any vertex from any other. So

$$\frac{v_m}{\psi_m} = \frac{v_n}{\psi_n}, \tag{3.151}$$

for any choice of m and n in $1 \ldots 2^N$. Thus $|v\rangle \propto |\psi\rangle$, and the state $|\psi\rangle$ is the unique state orthogonal to the set of states $\{|\phi_{jk}\rangle\}$. □

As for scaling of the number of copies, n, with the number of qubits, N, we can write the maximum probability of accepting a state orthogonal to $|\psi\rangle$ as

$$q = \max_{\substack{|\phi\rangle \\ \langle\phi|\psi\rangle = 0}} \langle\phi| \left[\frac{1}{N2^{N-1}} \sum_{k=1}^{N} \sum_{j=1}^{2^{N-1}} (\mathbb{1} - |\phi_{jk}\rangle\langle\phi_{jk}|) \right] |\phi\rangle \tag{3.152}$$

$$= \max_{\substack{|\phi\rangle \\ \langle\phi|\psi\rangle = 0}} \left[1 - \frac{1}{N2^{N-1}} \sum_{k=1}^{N} \sum_{j=1}^{2^{N-1}} |\langle\phi|\phi_{jk}\rangle|^2 \right] \tag{3.153}$$

$$= 1 - \frac{1}{N2^{N-1}} \min_{\substack{|\phi\rangle \\ \langle\phi|\psi\rangle = 0}} \sum_{k=1}^{N} \sum_{j=1}^{2^{N-1}} |\langle\phi|\phi_{jk}\rangle|^2 \tag{3.154}$$

$$\coloneqq 1 - \frac{f(N)}{N2^{N-1}}, \tag{3.155}$$

or that

$$n = \frac{N2^{N-1}}{f(N)} \cdot \frac{1}{\epsilon} \log \frac{1}{\delta}. \tag{3.156}$$

The summation in Eq. 3.155, and the quantity $f(N)$, is closely related to the *frame potential* for the set of states $|\phi_{jk}\rangle$ [16]. We can give a straightforward upper bound for $f(N)$. Note that for some element $|\phi_{rs}\rangle$,

$$f(N) = \min_{|\phi\rangle} \sum_{j,k} |\langle\phi|\phi_{jk}\rangle|^2 \leq \sum_{j,k} |\langle\phi_{rs}|\phi_{jk}\rangle|^2. \tag{3.157}$$

The index k indexes a block of N orthogonal states, and so this overlap vanishes unless $k = s$. As for the index j, in the worst case each N-state block will completely

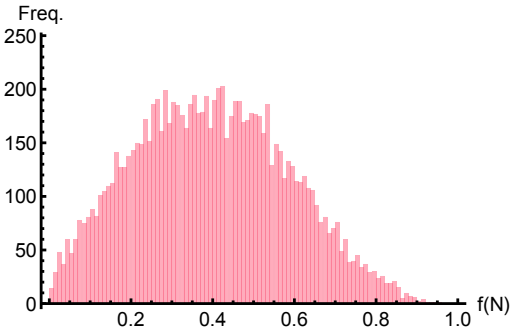

Figure 3.6: A histogram of the "frame potential" $f(N)$ for 10000 randomly chosen three-qubit states.

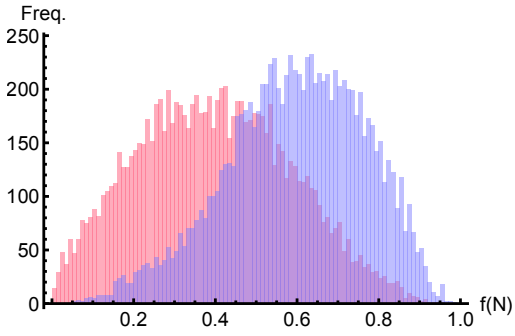

Figure 3.7: A histogram comparing the impact on the "frame potential" $f(N)$ given a verifier that is allowed to locally rotate their ansatz by a unitary $U = \hat{R}_x(\phi) \otimes \mathbb{1} \otimes \mathbb{1}$.

overlap with all the other blocks, and so the overlap picks up a factor of 1 for all 2^{N-1} blocks. Thus, $f(N) \leq 2^{N-1}$. This gives a lower bound for the copy complexity:

$$n \geq \frac{N}{\epsilon} \log \frac{1}{\delta}. \tag{3.158}$$

As for upper bounds, we would like the scaling in Eq. 3.156 to be as benign as possible; if we can show that $f(N)$ is either bounded from below by a constant, or even that it grows with N, this will yield an upper bound on the number of copies that grows more slowly than $N2^{N-1}$ and so this arbitrary state ansatz will outperform the scaling for direct fidelity estimation in terms of both ϵ and N. We explored numerically the value of the quantity $f(N)$ for 10000 randomly chosen three-qubit states, by constructing the states $|\phi_{jk}\rangle$ for the random state and then optimising over the adversary's state, $|\phi\rangle$. The results are shown in the histogram in Fig. 3.6. It is clear that our hope is not fulfilled, even for this simple case - while we know that the ansatz corresponds to a valid verification protocol from Theorem 24, and so the required number of copies of the state must remain finite, there are states where the quantity $f(N)$ can be made arbitrarily small. For these states, the number of required copies can be made arbitrarily large (albeit not infinite).

One procedure that a verifier might take to save this ansatz is to notice that any local verification strategy can be conjugated by local measurements, and remain a local verification strategy. Hence a verifier can take a target state $|\psi\rangle$ with a small $f(N)$, conjugate it by a local unitary $U|\psi\rangle = |\psi'\rangle$, derive the verification strategy via the ansatz for $|\psi'\rangle$, and then conjugate this strategy by U. The resultant strategy is guaranteed to be a valid verification strategy for $|\psi\rangle$, and has a different value for the frame potential f, in general. To explore the effect of this procedure, for the same

set of randomly chosen three-qubit states above, we can optimise over a verifier's choice of a specifically chosen local unitary U, designed to maximise the quantity f. We explored the effect of this procedure in a simple case, where the unitary U was chosen to be $U = \hat{R}_x(\phi) \otimes \mathbb{1} \otimes \mathbb{1}$ for computational expediency. The outcome of the verifier optimising over the parameter ϕ is shown in Fig. 3.7. In this case, this additional optimisation for the verifier is enough to pull f away from zero, for almost all randomly chosen three-qubit states. However, it is unclear whether this additional step can be folded into the ansatz in a straightforward and transparent manner, and that analytic bounds can then be derived.

3.8 Outlook

We have shown that, if one is sufficiently restrictive about the output of the verification protocol, that quantum state verification can in principle be much more efficient than characterising the state completely. For the states we have considered, the advantage is quadratic in the infidelity ϵ, and for stabilizer states, also exponential in the number of qubits N. This advantage can be enormous in practice. On the other hand, the proofs for deriving this advantage relied on one of two things: for two-qubit states, that we could efficiently elucidate the most general verification strategy; and for stabilizer states, that we could exploit the high degree of symmetry of states of this type. Both of these approaches seem problematic for deriving efficient verification strategies for arbitrary states. Additionally, the ansatz for arbitrary pure states that we have considered requires scrutiny if a provable advantage in copy complexity is to be derived. We have given preliminary numerical evidence that the frame potential can possibly be bounded away from zero, and that the copy complexity can then be bounded from above. An analytic proof of this would be the first step towards an efficient ansatz for arbitrary state verification. On the other hand, when the state to be verified, $|\psi\rangle$, is restricted to be a stabilizer state, the ansatz reduces to the strategy constructed from *stabilizer generators*, and not the more efficient strategy constructed from the full set of linearly-independent stabilizers. Hence there is some potential leeway to construct more efficient ansätze for arbitrary state verification.

One may also consider more relaxed frameworks for quantum state verification, in the same spirit as the "soft" verification scenario. For example, it may be reasonable to restrict the action of the adversary, given some physically-motivated context. Alternatively, it may be more analytically tractable to pursue average-case hardness results for arbitrary pure states, rather than reasoning about every possible target state $|\psi\rangle$.

CHAPTER 4

DIRECT FIDELITY ESTIMATION OF QUANTUM STATES ON A PHOTONIC CHIP

While the setting for state verification in Chapter 3 is perfectly valid, it has the drawback that, even in the "soft" scenario, the verifier gains very little information about the output state if the test fails (besides the information that the state is far from the target state $|\psi\rangle$). Fidelity estimation protocols, like those in § 2.1.4, try to strike a balance between information gleaned from the experiment, and the efficiency (in terms of copy complexity) of the protocol. Our aim in this chapter is to demonstrate that, under some reasonable assumptions, the optimal verification strategies in Chapter 3 are also optimal fidelity estimators. We demonstrate this analytically, but also produce a case-study of this fidelity estimation in practice, on output from a two-qubit silicon photonic chip.

We first give the theoretical description of the fidelity estimation protocols that we consider, in § 4.1. We then give a brief description of the photonic chip in § 4.2, and an analysis of a set of output data in § 4.3. We conclude in § 4.4 with a discussion of a popular technique for estimation in quantum photonics.

4.1 Fidelity estimation protocols

We first discuss the theoretical basis of fidelity estimation protocols. We start in § 4.1.1 by outlining a statement of the scenario in which we will operate, and list the set of requirements that we place on the protocol a priori. We then construct a family of estimators for the fidelity that satisfy these requirements in § 4.1.2, before showing in § 4.1.3 that, for two-qubit states, the protocol that we derived in the context of state verification is the minimum variance estimator within this family. We derive error bars for these estimates in § 4.1.4, and generalise them to account for "count miscalibration" in § 4.1.5.

4.1.1 Premise

Consider a verifier with access to an experiment that claims to produce copies of a target state $|\psi\rangle$. The verifier is given access to n copies of the state produced by the experiment, and must estimate the fidelity, F, of the output state with $|\psi\rangle$. We consider two scenarios, labelled X and Y:

Scenario X. *For the i^{th} trial, $0 \leq i \leq n$, the output of the experiment is a state ρ_{ik} drawn from an ensemble of states and associated probabilities indexed by k, $\{\alpha_{ik}, \rho_{ik}\}$, such that the expected fidelity for each trial is F: $\sum_k \alpha_{ik} \operatorname{tr}(|\psi\rangle\langle\psi|\rho_{ik}) = F$, for all i. For each trial the verifier randomly draws a projective measurement to apply indexed by j, P_{ij}, with probability μ_{ij}. The state ρ_{ik} is adversarially chosen such that the verifier gains the least information from the trial. As noted in § 3.2, in scenario X the verifier must maximise the information they receive for all i and so, assuming that each ρ_{ik} is chosen without reference to previous choices, the verifier's choice of strategy must be independent of i. As such, we drop the i index for α_{ik} and ρ_{ik}. Moreover, the adversary has no knowledge of the actual choice of measurement P_{ij}, and so must draw from an ensemble that is independent of i. As such, we also drop the i index for these quantities.*

Scenario Y. *The verifier deterministically divides the n total trials into blocks of size $n\mu_j$, for which measurement P_j is performed on each copy. The output of the experiment for all trials is assumed to be the same state ρ such that $\operatorname{tr}(|\psi\rangle\langle\psi|\rho) = F$, where ρ is adversarially chosen such that the verifier gains the least information from the trials, in aggregate.*

In both of the above scenarios, we impose the following properties on the strategy $\mathcal{S} = \{\mu_j, P_j\}$:

1. The estimate of the fidelity is constructed only from the sample mean of measurement outcomes, rather than any other more complicated statistic; i.e. the fidelity is constructed in scenario Y by estimating $\operatorname{tr}(\Omega\rho) = \sum_j \mu_j \operatorname{tr}(P_j\rho)$. In scenario X, the sample mean is $\operatorname{tr}(\Omega\rho) = \sum_{j,k} \mu_j\alpha_k \operatorname{tr}(P_j\rho_k) = \sum_j \mu_j \operatorname{tr}(P_j\bar{\rho})$, for the ensemble-averaged state $\bar{\rho} = \sum_k \alpha_k\rho_k$. Hence we can, without loss of generality, assume that the output of the experiment is the same adversarially-chosen state ρ for each trial.

2. The fidelity estimate should be consistent, independent of ρ. Specifically, for any state ρ such that $\operatorname{tr}(|\psi\rangle\langle\psi|\rho) = F$, the protocol should give the same expected sample mean, and hence the same expected estimate for the fidelity.

3. If the produced state ρ is exactly the target state $|\psi\rangle$, then the protocol must report $F = 1$.

4. The projectors P_j correspond to local observables (in both senses: P_j cannot be a non-local projector acting on a non-local part of a single ρ, and it cannot be a non-local projector acting across multiple copies of ρ).

5. Each P_j is a two-outcome projective measurement.

The spirit of Requirements 2 and 3 is similar to that of the *future-proof* requirement in § 3.2.1; we would like that an experimentalist can apply this protocol with no a priori knowledge about the output state, and be guaranteed a consistent estimate and uniform performance. Regarding Requirement 5, it may be the case in practice that even if a two-outcome measurement is performed, some "null" outcome occurs corresponding to the qubit being lost, or the measurement failing to operate. If this occurs, we do not consider a trial as having taken place.

We will label the random variable denoting the number of "+1" or "success" outcomes after n trials in each scenario as X and Y, respectively. In scenario X, the probability of recording a ± 1 outcome for a single trial is a mixture of Bernoulli variables, with weights μ_j:

$$\Pr[\pm 1 \text{ for single trial}] = \sum_j \mu_j \operatorname{tr}(P_j\rho)^k [1-\operatorname{tr}(P_j\rho)]^{1-k},\ k \in \{0,1\}. \tag{4.1}$$

Then the probability of recording k "+1" outcomes after n trials is the compound distribution

$$\Pr(X=k) = \binom{n}{k}\left(\sum_j \mu_j \operatorname{tr}(P_j\rho)\right)^k \left[1-\sum_j \mu_j \operatorname{tr}(P_j\rho)\right]^{n-k} = B(k;n,\sum_j \mu_j \operatorname{tr}(P_j\rho)). \tag{4.2}$$

In scenario Y, each block of $n\mu_j$ trials has a binomial distribution $Y_j \sim B(k;n\mu_j,\operatorname{tr}(P_j\rho))$. If we let the index j run from $1\ldots J$, we then have that $Y = \sum_{j=1}^J Y_j$. The probability of seeing k successes, $\Pr(Y=k)$, is then given by the convolution of the distributions governing Y_j:

$$\Pr(Y=k) = \sum_{\ell_2,\ell_3\ldots\ell_J=1}^{k} \left[B\left(k-\sum_{j=2}^J \ell_j; n\mu_1, \operatorname{tr}(P_1\rho)\right) \prod_{j=2}^J B\left(\ell_j; n\mu_j, \operatorname{tr}(P_j\rho)\right)\right]. \tag{4.3}$$

A key point to note is that both distributions have the same mean:

$$\mathbb{E}(X) = n\sum_j \mu_j \operatorname{tr}(P_j\rho), \tag{4.4}$$

and

$$\mathbb{E}(Y) = \mathbb{E}(\sum_j Y_j) = \sum_j \mathbb{E}(Y_j) = n\sum_j \mu_j \operatorname{tr}(P_j\rho). \tag{4.5}$$

Thus given that the fidelity estimate is constructed from the sample mean, the verifier expects the same point estimate for the fidelity in both scenarios. However, any quantity derived from other statistical moments may differ between scenarios X and Y.

4.1.2 Estimators for the fidelity

Consider a fidelity estimation strategy $\Omega = \sum_j \mu_j P_j$. The consequence of Requirement 3 is that each projector $P_j = |\psi\rangle\langle\psi| + \bar{P}_j$, for some residual projector $\bar{P}_j$. Thus we know that $|\psi\rangle$ is an eigenstate of Ω with eigenvalue $+1$; i.e. we can write Ω in its eigenbasis as $\Omega = |\psi\rangle\langle\psi| + \sum_\ell q_\ell |\phi_\ell\rangle\langle\phi_\ell|$. Then the expectation of Ω with respect to some state ρ is

$$\mathrm{tr}(\Omega\rho) = \mathrm{tr}(|\psi\rangle\langle\psi|\rho) + \sum_\ell q_\ell \,\mathrm{tr}(|\phi_\ell\rangle\langle\phi_\ell|\rho) = F + \sum_\ell q_\ell \,\mathrm{tr}(|\phi_\ell\rangle\langle\phi_\ell|\rho). \tag{4.6}$$

Now, for the most general set of parameters q_ℓ it is possible to vary the state ρ in this expression and change the sample mean, even if F is fixed. This is in contradiction with Requirement 2. There is only one viable choice for the verifier given this requirement, which is to set $q_\ell = q$, $\forall \ell$, i.e. when Ω is of the form $\Omega = |\psi\rangle\langle\psi| + q\sum_\ell |\phi_\ell\rangle\langle\phi_\ell| = (1-q)|\psi\rangle\langle\psi| + q\mathbb{1}$. Then

$$F = \frac{1}{1-q}\left[\mathrm{tr}(\Omega\rho) - q\right]. \tag{4.7}$$

We will denote X_F and Y_F the experimental estimate of the fidelity in scenarios X and Y, respectively. In analogy with Eq. 4.7, we construct our estimate for the fidelity in scenario X as

$$X_F = \frac{1}{1-q}\left[\frac{X}{n} - q\right], \tag{4.8}$$

and equivalently for Y_F. One may worry about the potential bounding values for the variables X_F (and Y_F). It is clear that, as $X \leq n$, the estimate $X_F \leq 1$. However, if $X < nq$ then the fidelity estimate X_F can be negative. This ceases to be a problem in the limit of large n; the weak law of large numbers implies that the sample mean will converge in probability to the expected value, and the expected value of X is

$$\mathbb{E}(X) = n\,\mathrm{tr}(\Omega\rho) = n[q + F(1-q)] \geq nq. \tag{4.9}$$

However, it is instructive to calculate the regime in which we expect to see negative estimates for the fidelity, given a particular choice of n, F and q. The probability that $X \leq nq$ can be bounded by the concentration inequality

$$\Pr(X \leq nq) \leq \exp\{-nD(q\|q + F(1-q))\}, \tag{4.10}$$

where $D(\cdot\|\cdot)$ is the relative entropy. Suppose we tolerate some fixed probability of a negative estimate for the fidelity, i.e. $\Pr(X \leq nq) = \delta$. This expression implies that, for a verifier capable of carrying out n trials and using a strategy with parameter q, there is a lower bound to F above which we will report a negative fidelity with probability bigger than δ. In the two-qubit experiments that follow, $q \leq 0.6$ and $n > 1000$. We

take as a worst-case that the output state is the maximally mixed state; and so we have a lower bound on the fidelity $F \geq 0.25$. Given these parameters, $\Pr(X \leq nq) \lesssim 10^{-10}$ and so we need not worry about negative fidelity estimates for the case-study below. However, the situation may be more problematic for worse strategies (those with higher q), fewer trials n, or for estimates of states with particularly low fidelity ($F \approx 0$).

4.1.3 Estimator optimality

Given the family of possible measurement strategies Ω satisfying the strategy requirements listed above, we would like to choose the strategy that produces the highest quality estimate of F. We now show the following:

Lemma 25. *Given a family of measurement strategies $\Omega = \sum_j \mu_j P_j = (1-q)|\psi\rangle\langle\psi| + q\mathbb{1}$, for varying q, the variance of the estimate of the fidelity, F, in both scenarios X and Y is in the worst case*

$$Var(Y_F) \leq \frac{1-F}{n}\left(F + \frac{q}{1-q}\right) = Var(X_F). \tag{4.11}$$

Moreover, there exist some plausible choices of Ω that saturate this inequality, and hence the minimum variance estimator in either scenario in the worst case is the strategy that minimises q.

Proof. Scenario X - X is binomially distributed according to $B(n, \sum_j \mu_j \operatorname{tr}(P_j\rho))$, and so

$$\begin{aligned}
\operatorname{Var}(X) &= n\left[\sum_j \mu_j \operatorname{tr}(P_j\rho)\right]\left[1 - \sum_j \mu_j \operatorname{tr}(P_j\rho)\right] &(4.12)\\
&= n\operatorname{tr}(\Omega\rho)(1-\operatorname{tr}(\Omega\rho)) \\
&= n[q + (1-q)F][1 - q - (1-q)F] \\
&= n[q(1-q) - Fq(1-q) + F(1-q)^2 - F^2(1-q)^2] \\
&= n(1-q)[q + F(1-2q) - F^2(1-q)]. &(4.13)
\end{aligned}$$

The variance of X_F is just

$$\begin{aligned}
\operatorname{Var}(X_F) = \frac{1}{n^2}\cdot\frac{1}{(1-q)^2}\operatorname{Var}(X) &= \frac{1}{n}\left(\frac{q + F(1-2q) - F^2(1-q)}{1-q}\right) \\
&= \frac{1-F}{n}\left(F + \frac{q}{1-q}\right), &(4.14)
\end{aligned}$$

and so the variance is independent of the adversary's choice of ρ.

Scenario Y - For each block of $n\mu_j$ measurements, each setting j is associated with a binomially-distributed random variable $Y_j \sim B(n\mu_j, \operatorname{tr}(P_j\rho))$. The random

variable Y governing estimates of $\mathrm{tr}(\Omega\rho)$ given n trials is then the composite variable $Y = \sum_j Y_j$. Given uncorrelated samples, the variance of Y is $\mathrm{Var}(Y) = \mathrm{Var}(\sum_j Y_j) = \sum_j \mathrm{Var}(Y_j) = \sum_j n\mu_j \mathrm{tr}(P_j\rho)(1-\mathrm{tr}(P_j\rho))$. Given that each measurement accepts the target with certainty, we can expand the projector $P_j = |\psi\rangle\langle\psi| + \bar{P}_j$ for some residual projector $\bar{P}_j$. Then we can rewrite $\mathrm{Var}(Y)$ as

$$\begin{aligned}
\mathrm{Var}(Y) &= n\sum_j \mu_j \mathrm{tr}(P_j\rho)(1-\mathrm{tr}(P_j\rho)) && (4.15)\\
&= n\sum_j \mu_j \mathrm{tr}[(|\psi\rangle\langle\psi| + \bar{P}_j)\rho]\{1-\mathrm{tr}[(|\psi\rangle\langle\psi| + \bar{P}_j)\rho]\} && (4.16)\\
&= n\sum_j \mu_j [F + \mathrm{tr}(\bar{P}_j\rho)]\{1-[F+\mathrm{tr}(\bar{P}_j\rho)]\} && (4.17)\\
&= n\sum_j \mu_j [F + \mathrm{tr}(\bar{P}_j\rho) - F^2 - 2F\,\mathrm{tr}(\bar{P}_j\rho) - \mathrm{tr}(\bar{P}_j\rho)^2] && (4.18)\\
&= n[F(1-F) + (1-2F)\sum_j \mu_j \mathrm{tr}(\bar{P}_j\rho) - \sum_j \mu_j \mathrm{tr}(\bar{P}_j\rho)^2]. && (4.19)
\end{aligned}$$

Now, $\sum_j \mu_j \bar{P}_j = q(\mathbb{1} - |\psi\rangle\langle\psi|)$, and so $\sum_j \mu_j \mathrm{tr}(\bar{P}_j\rho) = q(1-F)$. So

$$\mathrm{Var}(Y) = n\left[F(1-F) + q(1-F)(1-2F) - \sum_j \mu_j \mathrm{tr}(\bar{P}_j\rho)^2\right]. \quad (4.20)$$

Using Jensen's inequality and the fact that x^2 is convex implies that

$$\sum_j \mu_j \mathrm{tr}(\bar{P}_j\rho)^2 \geq \left(\sum_j \mu_j \mathrm{tr}(\bar{P}_j\rho)\right)^2 = q^2(1-F)^2, \quad (4.21)$$

and so we can bound the variance as

$$\begin{aligned}
\mathrm{Var}(Y) &\leq n[F(1-F) + q(1-F)(1-2F) - q^2(1-F)^2]\\
&= n\left[(1-q)\left[F(1-F) + q(1-F)^2\right]\right]. && (4.22)
\end{aligned}$$

Then, the variance of our estimate for the fidelity in scenario Y is given by $\mathrm{Var}(Y_F) = \frac{1}{n^2(1-q)^2}\mathrm{Var}(Y)$, or that

$$\mathrm{Var}(Y_F) \leq \frac{1-F}{n}\left(F + \frac{q}{1-q}\right) = \mathrm{Var}(X_F). \quad (4.23)$$

It is in the interest of an adversary to choose ρ such that the variance is as large as possible; i.e. to saturate the bound in Eq. 4.21. This is saturated when each term $\mathrm{tr}(\bar{P}_j\rho)$ is equal, for all j. There are valid choices of $\{\mu_j, P_j\}$ where this is possible; for example, if $\mathrm{tr}(P_j)$ is a constant independent of j, then the adversary can pick

$$\rho = \left(F - \frac{1-F}{d-1}\right)|\psi\rangle\langle\psi| + \frac{1-F}{d-1}\mathbb{1} \quad (4.24)$$

and so $\mathrm{tr}(\bar{P}_j\rho) = \frac{1-F}{d-1}\mathrm{tr}(\bar{P}_j)$, which is independent of j. Hence in the worst case, there exists an Ω such that $\mathrm{Var}(Y_F) = \mathrm{Var}(X_F)$. □

This optimisation is precisely the one carried out in the previous chapter, for two-qubit states in particular. Therefore we have the following corollary:

Corollary 26. *The minimum-variance local estimator $\Omega = \sum_j \mu_j P_j$ for the fidelity to the target state $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$ for the range $0 < \theta < \frac{\pi}{2}$, $\theta \neq \frac{\pi}{4}$, is that given in Theorem 18:*

$$\Omega = \frac{2-\sin(2\theta)}{4+\sin(2\theta)} P^{+}_{ZZ} + \frac{2(1+\sin(2\theta))}{3(4+\sin(2\theta))} \sum_{k=1}^{3} (\mathbb{1} - |\phi_k\rangle\langle\phi_k|), \tag{4.25}$$

where the states $|\phi_k\rangle$ are

$$|\phi_1\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{4.26}$$

$$|\phi_2\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\cot\theta}}|1\rangle\right), \tag{4.27}$$

$$|\phi_3\rangle = \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle + \frac{1}{\sqrt{1+\cot\theta}}|1\rangle\right) \otimes \left(\frac{1}{\sqrt{1+\tan\theta}}|0\rangle - \frac{1}{\sqrt{1+\cot\theta}}|1\rangle\right). \tag{4.28}$$

This estimator has variance

$$Var(X_F) = Var(Y_F) = \frac{1-F}{n}(F + 1 + \sin\theta\cos\theta). \tag{4.29}$$

4.1.4 Error bars

Denote the experimental estimate of the fidelity as $\tilde{F}$ (where in the experiments defined here, either $\tilde{F} = X_F$ or $\tilde{F} = Y_F$, depending on the scenario), and the true value as F. Then to derive a (possibly lop-sided) confidence interval on the estimate of F, $[\tilde{F} + t^{+}_{F}, \tilde{F} - t^{-}_{F}]$, one must invert the expressions

$$\Pr[F - \tilde{F} \geq t^{+}_{F}] \leq \delta \tag{4.30}$$

$$\Pr[F - \tilde{F} \leq t^{-}_{F}] \leq \delta \tag{4.31}$$

for t^{+}_{F} and t^{-}_{F}, where $1-\delta$ is some prespecified confidence. There are three ways that one might contemplate doing this: (i) make full use of the mathematical description of the distributions of X_F and Y_F to derive an exact confidence interval; (ii) use some asymptotic approximation of the exact distribution to derive an approximate confidence interval; (iii) make use of an appropriate concentration inequality that only relies on some partial information about the true distribution.

Consider first the strategy $\Omega = \sum_j \mu_j P_j$ in scenario X; i.e. for each copy of ρ, the verifier picks a setting indexed by j with probability μ_j. In this scenario the probability of outputting outcome "success" for k of n trials has binomial distribution $\Pr(X = k) = B(k; n, \sum_j \mu_j \mathrm{tr}(P_j\rho))$. In this instance, we can take approach (i), and

derive an exact binomial proportion confidence interval, or "Clopper-Pearson" interval [63]. Denote the true mean $\frac{\mathbb{E}(X)}{n}$, and the experimentally-measured sample mean as $\frac{X}{n}$. We would like to make a statement that the true mean is close to the sample mean, with some statistical confidence, i.e. we require that

$$\Pr[X \geq \mathbb{E}(X) + nt^-] \leq \delta \;\&\; \Pr\left[X \leq \mathbb{E}(X) - nt^+\right] \leq \delta, \tag{4.32}$$

or by inserting the distribution for X as

$$\Pr\left[\frac{1}{n}B(k;n,\sum_j \mu_j \operatorname{tr}(P_j\rho)) \geq \sum_j \mu_j \operatorname{tr}(P_j\rho) + t^-\right] \leq \delta \tag{4.33}$$

$$\&\; \Pr\left[\frac{1}{n}B(k;n,\sum_j \mu_j \operatorname{tr}(P_j\rho)) \leq \sum_j \mu_j \operatorname{tr}(P_j\rho) - t^+\right] \leq \delta. \tag{4.34}$$

In principle, this can then be explicitly solved for t^+ and t^-. The original solution is given by Clopper and Pearson in [63]. The resultant error bars are given by

$$t^- \geq X - I_X\left(\frac{\delta}{2}; nX, n - nX + 1\right); \tag{4.35}$$

$$t^+ \geq I_X\left(1 - \frac{\delta}{2}; nX + 1, n(1-X)\right) - X. \tag{4.36}$$

where $I_X(a;b,c)$ is the a^{th} quantile of the "incomplete" Beta distribution for random variable X, with "shape" parameters b and c. Given the skewness of the binomial distribution, we do not expect $t^+ = t^-$ in general. Given the simple linear relationship between the sample mean and fidelity given by Eq. 4.7, we can denote the experimental estimate of the fidelity as $X_F = \frac{1}{1-q}\left(\frac{X}{n} - q\right)$ and the true fidelity as $F = \frac{1}{1-q}\left(\frac{\mathbb{E}(X)}{n} - q\right)$, respectively. Then taking the inequalities in Eq. 4.32 and applying the same linear transformation gives

$$\Pr\left[X_F \geq F + \frac{1}{1-q}(t^- - q)\right] \leq \delta \;\&\; \Pr\left[X_F \leq F - \frac{1}{1-q}\left(t^+ - q\right)\right] \leq \delta, \tag{4.37}$$

This implies that we can bound the estimate of the fidelity X_F in the interval $[X_F + t_F^+, X_F - t_F^-]$ where

$$t_F^- = \frac{1}{1-q}(t^- - q) \geq X_F - \frac{1}{1-q}(I_X(\delta; nX, n - nX + 1) - q) \tag{4.38}$$

$$t_F^+ = \frac{1}{1-q}(t^+ - q) \geq \frac{1}{1-q}(I_X(1-\delta; nX + 1, n(1-X)) - q) - X_F. \tag{4.39}$$

Alternatively, one may try to find a simpler confidence interval by appealing to (ii); e.g. by considering the normal approximation to the binomial distribution. However, these approximations are only completely faithful to the underlying distribution in the limit of a large number of trials, and tend to perform poorly when the number of

trials is small or the trial bias is close to 0 or 1. We would like to operate in a scenario where an experimentalist can run the fidelity estimation protocol for any period of time, and be guaranteed a point estimate with a well-defined error bar; in particular, we would like the error bars to be meaningful even for small n. We cannot guarantee the correctness of approximate intervals constructed by these approximations, in this case. Alternatively, one may try (iii), and make use of appropriate concentration bounds for X. However, these are guaranteed to produce a looser confidence interval than the exact solution, as they rely on partial information about the experiment.

In scenario Y, we no longer have the case that the underlying distribution is a simple binomial; instead, we have that the samples are drawn from a convolution of distributions, with each element given by a binomial distribution and weights given by the parameters μ_j. In this case, we cannot proceed with approach (i) as above and extract an analytic closed form for the exact confidence interval. Moreover, we also cannot appeal to (ii) for the reasons stated above. We are left with approach (iii), i.e. applying an appropriate concentration inequality that relies on some partial information about Y. In this case, for example, we can make use of the Chernoff-Hoeffding inequalities:

$$\Pr[Y_F \geq F + t_F^-] \leq \exp\left\{-nD\left(F + t_F^- \| F\right)\right\} \leq \delta, \tag{4.40}$$

$$\Pr[Y_F \leq F - t_F^+] \leq \exp\left\{-nD\left(F - t_F^+ \| F\right)\right\} \leq \delta, \tag{4.41}$$

for some pre-specified confidence $1-\delta$. As a reminder to the reader, $D(\cdot\|\cdot)$ is the relative entropy, which for Bernoulli variables a and b we can expand as:

$$D(a\|b) = a\ln\frac{a}{b} + (1-a)\ln\frac{1-a}{1-b}. \tag{4.42}$$

The goal is to extract the behaviour of t_F^+ and t_F^-, the bounds of the error bar, as a function of n, F and δ. In general, the expression for the relative entropy is not analytically invertible and the error bars must be extracted numerically. However, as in § 3.2.3, we can write down a couple of useful edge cases: (a) in the best possible case, $F = 1$; then

$$D(F - t\|F) \to \ln\frac{1}{1-t}. \tag{4.43}$$

Substituting this into the Chernoff bound gives

$$t = 1 - \left(\frac{1}{\delta}\right)^{-\frac{1}{n}} \approx \frac{1}{n}\ln\frac{1}{\delta}; \tag{4.44}$$

(b) in the worst case, $F \neq 1$ and t is small. Taking a Taylor series expansion to first order for small t gives

$$D \to \frac{t^2}{2F(1-F)}. \tag{4.45}$$

Substituting into the Chernoff bound yields

$$t = \sqrt{\frac{2F(1-F)}{n} \ln \frac{1}{\delta}}. \tag{4.46}$$

Note that this is only better than the bound given by the weaker Hoeffding's inequality by a constant factor; however, this difference might be useful in practice. Additionally, the confidence interval given by the Chernoff-Hoeffding inequality is asymptotically optimal (see § 2.1.6); however, for fixed sample sizes this interval may be looser than in the exact case.

4.1.5 Count miscalibration

As we will observe in the experimental case study in § 4.2 onward, it will often be the case that, even if the measurement prescription requires $n\mu_j$ copies of the state ρ to be measured with P_j, this may not be exactly achievable in practice. While this may be achievable in implementations with static qubits, quantum optics is prone to (a) photon loss; (b) variation in counts due to the probabilistic nature of single-photon sources; and (c) fluctuations in the number statistics of quantum states in the experiment. All of these factors lead to discrepancies between the number of measurements performed with respect to a particular setting, and the number of measurements specified by the strategy. We show below that the effect of this miscalibration is calculable. We focus on scenario Y, as this corresponds to the photonic implementation discussed below. While the proof of Theorem 27 looks somewhat involved, we note that it is almost identical to the original proof of the Chernoff-Hoeffding inequalities in Eq. 4.40 [114].

Theorem 27. *Consider a strategy $\Omega = \sum_j \mu_j P_j$, in scenario Y, i.e. where each measurement setting j should be applied deterministically to a block of $n\mu_j$ independent copies of ρ. Suppose that an experiment instead carries out measurements P_j on $n\alpha_j$ copies with a number of successes given by the binomial random variable Y_j, for $\sum_j \alpha_j = 1$. Then let $w_j = \frac{\mu_j}{\alpha_j}$ and $w = \frac{1}{J}\sum_{j=1}^{J} w_j$. Then the verifier can construct an estimator $Y = \sum_j w_j Y_j$, with error bars given by the Chernoff-Hoeffding inequalities*

$$Pr[Y \geq (p+t)n] \leq \exp\left\{-D\left(\frac{p+t}{w}\middle\|\frac{p}{w}\right)n\right\}, \tag{4.47}$$

$$Pr[Y \leq (p-t)n] \leq \exp\left\{-D\left(\frac{p-t}{w}\middle\|\frac{p}{w}\right)n\right\}. \tag{4.48}$$

Proof. In this scenario, for setting j the verifier can take data on $n\alpha_j$ copies and weight the measured number of successes by a factor $\frac{\mu_j}{\alpha_j}$, i.e. that $Y = \sum_j w_j Y_j$, for $w_j = \frac{\mu_j}{\alpha_j}$. Then $\mathbb{E}(Y) = \sum_j \frac{\mu_j}{\alpha_j} n\alpha_j \operatorname{tr}(P_j\rho) = n\sum_j \mu_j \operatorname{tr}(P_j\rho)$, so the expectation of the

output (i.e. the verifier's point estimate of the fidelity) is the same as if the experiment ran with perfect weights. As for error bars, label the i^{th} trial of setting j as Y_{ij}, where Y_{ij} is a Bernoulli trial with success probability $p_{ij} = \mathrm{tr}(P_j\rho)$. Then, $Y = \sum_{j=1}^{J} w_j \sum_{i=1}^{n\alpha_j} Y_{ij}$. Also, denote $\mathbb{E}(Y) = \sum_j \mu_j \mathrm{tr}(P_j\rho) \coloneqq \sum_j \mu_j p_j \coloneqq p$. Then

$$\Pr[Y \geq (p+t)n] = \Pr[e^{\lambda Y} \geq e^{(p+t)n}], \tag{4.49}$$

for some $\lambda > 0$, and so by Markov's inequality we have that

$$\Pr[Y \geq (p+t)n] \leq \frac{\mathbb{E}(e^{\lambda Y})}{e^{\lambda(p+t)n}}. \tag{4.50}$$

Then substituting in $Y = \sum_{ij} w_j Y_{ij}$ and assuming that each trial is independent gives

$$\Pr[Y \geq (p+t)n] \leq \frac{\prod_{ij} \mathbb{E}(e^{\lambda w_j Y_{ij}})}{e^{\lambda(p+t)n}}. \tag{4.51}$$

Now, $e^{\lambda r}$ is convex in r, so we know that, for $r \in [0, w]$, $e^{\lambda r} \leq 1 - \frac{r}{w} + \frac{r}{w} e^{\lambda w}$. Thus,

$$\Pr[Y \geq (p+t)n] \leq \frac{\prod_{ij} \mathbb{E}\left(1 - Y_{ij} + Y_{ij} e^{\lambda w_j}\right)}{e^{\lambda(p+t)n}}. \tag{4.52}$$

The expectation of Y_{ij} is $p_j = \mathrm{tr}(P_j\rho)$. So

$$\Pr[Y \geq (p+t)n] \leq \frac{\prod_{ij} \left(1 - p_j + p_j e^{\lambda w_j}\right)}{e^{\lambda(p+t)n}}. \tag{4.53}$$

Then, by the AM/GM inequality $\prod_k x_k \leq \left(\frac{1}{n}\sum_k x_k\right)^n$, we have that

$$\Pr[Y \geq (p+t)n] \leq \left(\frac{\frac{1}{n}\sum_{ij}\left(1 - p_j + p_j e^{\lambda w_j}\right)}{e^{\lambda(p+t)}}\right)^n. \tag{4.54}$$

Now, the weights w_j and probabilities of success p_j are independent of i, so the summation over i vanishes:

$$\Pr[Y \geq (p+t)n] \leq \left(\frac{\frac{1}{n}\sum_{ij}\left(1 - p_j + p_j e^{\lambda w_j}\right)}{e^{\lambda(p+t)}}\right)^n \tag{4.55}$$

$$= \left(\frac{1 - \sum_j \alpha_j p_j + \sum_j \alpha_j p_j e^{\lambda w_j}}{e^{\lambda(p+t)}}\right)^n \tag{4.56}$$

$$= \left(\frac{1 - \sum_j \frac{\alpha_j}{\mu_j} \mu_j p_j + \sum_j \frac{\alpha_j}{\mu_j} \mu_j p_j e^{\lambda w_j}}{e^{\lambda(p+t)}}\right)^n \tag{4.57}$$

$$\leq \left(\frac{1 - p\sum_j \frac{1}{w_j} + p\sum_j \frac{1}{w_j} e^{\lambda w_j}}{e^{\lambda(p+t)}}\right)^n. \tag{4.58}$$

Then, using Jensen's inequality and the fact that $\frac{1}{x}$ is convex, we have that $\sum_j \frac{1}{w_j} \geq \frac{1}{\sum_j w_j}$ and so we can simplify the concentration inequality as

$$\Pr[Y \geq (p+t)n] \leq \left(\frac{1 - \frac{pJ}{\sum_j w_j} + \frac{pJ}{\sum_j w_j} e^{\frac{\lambda}{J}\sum_j w_j}}{e^{\lambda(p+t)}}\right)^n = \left(\frac{1 - \frac{p}{w} + \frac{p}{w} e^{\lambda w}}{e^{\lambda(p+t)}}\right)^n, \tag{4.59}$$

where $w = \frac{1}{J}\sum_j w_j$. Now, this expression should hold for all λ, and as we are searching for the tightest bound, we must optimise over the choice of this parameter. Differentiating the expression (ignoring the exponent n) gives

$$\frac{d}{d\lambda}\left[\left(1-\frac{p}{w}\right)e^{-\lambda(p+t)}+\frac{p}{w}e^{\lambda(w-p-t)}\right] = -(p+t)\left(1-\frac{p}{w}\right)e^{-\lambda(p+t)}+(w-p-t)\frac{p}{w}e^{\lambda(w-p-t)}; \tag{4.60}$$

so setting to zero and rearranging for λ gives

$$\lambda = \frac{1}{w}\ln\left[\frac{(p+t)(w-p)}{p(w-p-t)}\right]. \tag{4.61}$$

All that remains it to substitute into Eq. 4.59. Taking instead the (natural) log of this expression, and ignoring the exponent n, we have that

$$\begin{aligned}
\ln\left[\left(1-\frac{p}{w}\right)e^{-\lambda(p+t)}+\frac{p}{w}e^{\lambda(w-p-t)}\right] &= \ln\left[e^{-\lambda(p+t)}\left(1-\frac{p}{w}+\frac{p}{w}e^{\lambda w}\right)\right] && (4.62)\\
&= -\lambda(p+t)+\ln\left(1-\frac{p}{w}+\frac{p}{w}e^{\lambda w}\right) && (4.63)\\
&= -\frac{p+t}{w}\ln\left[\frac{(p+t)(w-p)}{p(w-p-t)}\right] && (4.64)\\
&\quad +\ln\left(1-\frac{p}{w}+\frac{p}{w}\cdot\frac{(p+t)(w-p)}{p(w-p-t)}\right) && (4.65)\\
&= -\frac{p+t}{w}\ln\left[\frac{(p+t)(w-p)}{p(w-p-t)}\right]+\ln\left(\frac{w-p}{w-p-t}\right) && (4.66)\\
&= -\frac{p+t}{w}\ln\frac{p+t}{p}-\frac{p+t}{w}\ln\frac{w-p}{w-p-t}+\ln\frac{w-p}{w-p-t} && (4.67)\\
&= -\frac{p+t}{w}\ln\frac{p+t}{p}-\left(1-\frac{p+t}{w}\right)\ln\frac{w-p-t}{w-p} && (4.68)\\
&= -\frac{p+t}{w}\ln\frac{\frac{p+t}{w}}{\frac{p}{w}}-\left(1-\frac{p+t}{w}\right)\ln\frac{1-\frac{p+t}{w}}{1-\frac{p}{w}} && (4.69)\\
&=: -D\left(\frac{p+t}{w}\middle\|\frac{p}{w}\right). && (4.70)
\end{aligned}$$

In the case where $\mu_j = \alpha_j, \forall j$, we have that $w = 1$; and this reduces to the familiar expression for the relative entropy. Thus we have that

$$\Pr[Y \geq (p+t)n] \leq \exp\left\{-D\left(\frac{p+t}{w}\middle\|\frac{p}{w}\right)n\right\}. \tag{4.71}$$

An equivalent derivation holds for the lower bound on Y. □

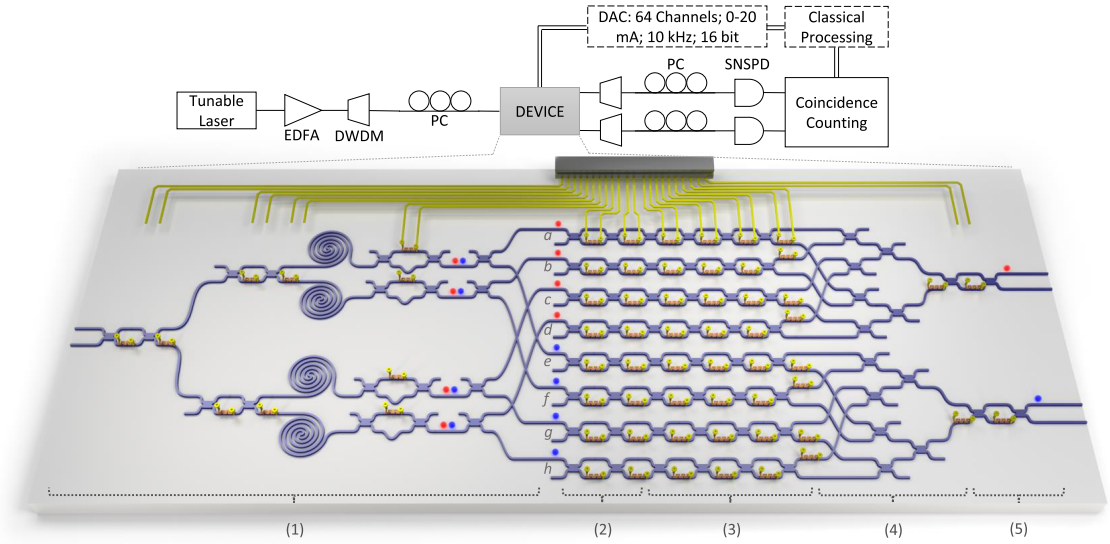

Figure 4.1: A schematic of the two-qubit photonic chip.

4.2 The photonic chip

The experimental setup that we use as a test bed is a two-qubit silicon photonics chip. In this section, we sketch the structure and operation of the chip; our treatment should not be considered comprehensive, and we refer the reader to [190] for a more thorough description. As a disclaimer, the fabrication and operation of this chip is not the novel aspect of this work - we include a discussion for the sake of completeness.

Fig. 4.1 is a schematic of the silicon photonic chip. Input laser light is generated off-chip, via a tunable, continuous wave (CW) laser. The light is then amplified with an optical fibre amplifier (EDFA) and the spectrum is filtered by a dense wavelength-division multiplexing (DWDM) module. The polarisation of this light is then controlled by in-line polarisation controllers, in order to optimise the coupling of light into the chip. The coupling from fibre into the chip is via a V-groove fibre array.

As for the chip itself, its footprint is 7.1 mm × 1.9 mm, and is built with readily available CMOS-based fabrication techniques. It contains a multitude of components: (i) four spiral-waveguide spontaneous four-wave mixing (SFWM) photon-pair sources; (ii) four laser pump rejection filters; (iii) 82 multi-mode interferometer (MMI) beam splitters; and (iv) 58 simultaneously-operational thermo-optic phase shifters.

The operation of the chip itself can be split into five sections, as labelled in Fig. 4.1. The function of each section is, broadly, the following: (1) conversion of laser light into entangled photon pairs, using the four spiral-waveguide SFWM sources;

(2) preparing initial single-qubit states, in an appropriate Fock basis; (3) implementing single-qubit operations; (4) implementing a two-qubit, non-local operation; (5) rotating the output state so that it can be measured in a desired basis. Parts (1), (3) and (4), as a whole, can in principle be used to implement any given $SU(4)$ operation.

Once the unitary and measurement basis has been specified, photons emerging from the device are collected in fibre by the same V-groove fibre array. Two analogous output DWDMs are used to separate the signal (red) and idler (blue) photons. Photons are detected by two fibre-coupled superconducting nanowire single-photon detectors (SNSPDs), with the polarisations of output photons again optimised by in-line polarisation controllers (PC). Coincidence counting logic records the two-photon coincidence events. Phase shifters on the device are configured through a digital-to-analog converter (DAC), controlled from a computer.

This design has a number of attractive features. It in principle performs universal two-qubit processing with high fidelity, as there is inherent phase stability of the optical paths and waveguide interferometric structures due to the monolithic construction of the device. All the thermal phase-shifters can be operated simultaneously, and can be reprogrammed at kHz rates. Experiments carried out on-chip are also repeatable under continuous operation, without recalibration.

4.3 Protocol implementation and analysis

The above fidelity estimation protocol was run on the photonic chip for input states of the form $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, for $\theta = \frac{k\pi}{32}$, $k = 0 \ldots 16$. While the projectors in the optimal strategy for states of this form in Theorem 18 have rank 2 and 3, the chip itself is only capable of measuring rank 1 projectors. Hence each measurement must be "unpacked" into a lower rank form; the precise details of the measurements, and the duration for which they were performed, are shown in Table B.1 in Appendix B. We should also note that for the Bell state $\theta = \frac{\pi}{4}$ we also used the strategy in Theorem 18, despite it not being exactly optimal in this instance. For each choice of k, the total integration time was fixed at 400 seconds, giving a total integration time for the whole data set of ~ 2 hours. Calculation of the fidelities and their associated error bars for the entire data set in Mathematica is computationally negligible, taking a few seconds. We take $1 - \delta = 0.95$.

The resultant fidelities are plotted in Fig 4.2. The size of these error bars depends on a handful of factors:

1. *Dependence on θ.* If the sample mean has error bars $t^{\pm}$ and the fidelity $t_F^{\pm}$, then we have seen that $t_F^{\pm} \sim \frac{1}{1-q(\theta)} t^{\pm} = (2 + \sin\theta\cos\theta)t^{\pm}$. Hence, if all other factors

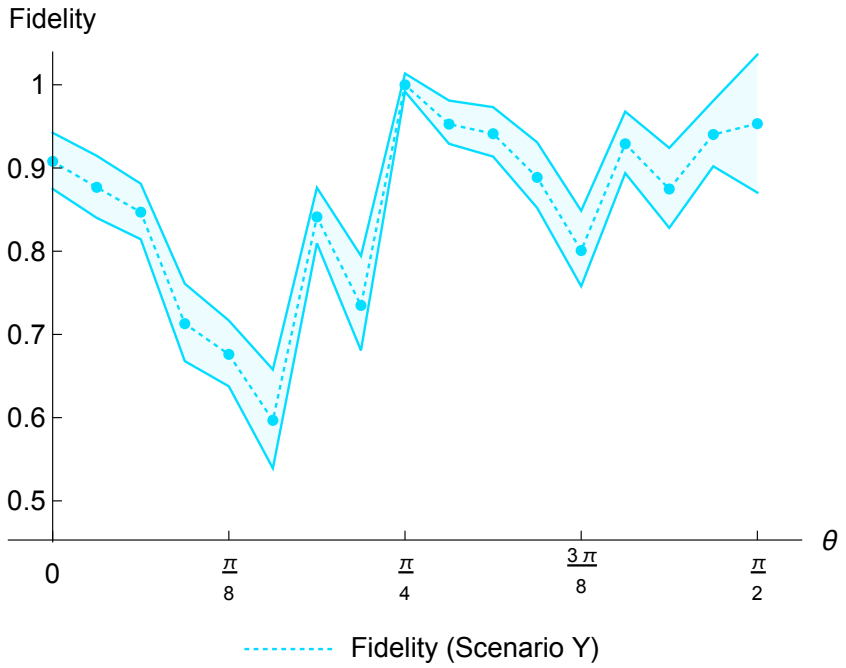

Figure 4.2: The fidelity estimate and confidence interval for target states of the form $\sin\theta|00\rangle + \cos\theta|11\rangle$. Point estimates extracted from output data from the chip are given by filled circles; with interpolating values given by the dotted line (included only as a visual aid). The shaded region bounded by solid lines denotes a 95% confidence interval derived from the Chernoff-Hoeffding error bars in Theorem 27.

are fixed, we expect smaller error bars for states close to product, and larger error bars for states close to the Bell state.

2. *Dependence on F.* The relative entropy is significantly greater, and hence the Chernoff-Hoeffding bounds in Thm. 27 yield a much smaller error bar for a fixed number of counts, if the fidelity is close to $F = 1$.

3. *Dependence on n.* The total number of measurements per choice of θ varies due to both stochastic mechanisms such as photon loss, and systematic mechanisms based on the operational structure of the chip.

4. *The effect of count miscalibration.* The amount of count miscalibration varies between choices of θ, in an unpredictable way.

5. *Choice of experimental scenario.* Operating in scenario Y, rather than scenario X, means that we cannot use the exact Clopper-Pearson error bar, which is generically smaller.

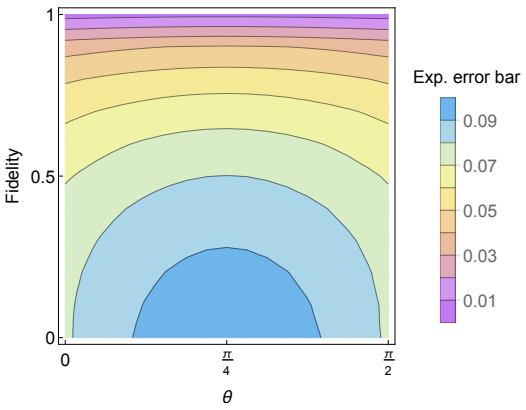

Figure 4.3: The expected size of the error bar, given a point estimate of the fidelity and a choice of θ, for $n = 1000$ and $\delta = 0.05$.

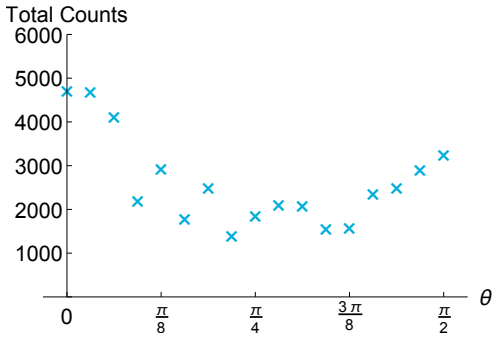

Figure 4.4: The total number of recorded counts after integrating each choice of θ for 400 seconds.

6. *Choice of concentration inequality.* Deriving error bars from a looser concentration inequality, such as Hoeffding's inequality, makes a notable difference on the size of the error bars.

We can be quantitative about the effect of each of these factors on the size of the confidence intervals in Fig. 4.2.

Regarding Point 1, we can see in Fig. 4.3 that designing an experiment that produces high-fidelity states, while being beneficial in an obvious sense, also yields a greater relative entropy and so is beneficial in that it becomes much quicker to verify (or, that it requires far fewer total measurements to verify to within a particular fixed confidence interval than a low-fidelity state). This would also be the case in scenario X; the size of the Clopper-Pearson error bars is also minimised for extremal biases (i.e. when the sample mean is expected to be 0 or 1).

Regarding Point 2, there is a penalty incurred in the size of the error bar as the target state becomes more entangled. This is because the fidelity is given by Eq. 4.7, where there is a conversion factor of $\frac{1}{1-q}$ in converting from the sample mean to the fidelity. If the verifier had access to the projector $|\psi\rangle\langle\psi|$, then $q = 0$, the sample mean would directly give an estimate of the fidelity, and there would be no penalty in converting between the two. However, the verifier is restricted to implementing local measurements; and so $q \neq 0$ and there is a θ-dependent penalty. Thus this effect can be interpreted as the penalty for restricting to local measurements. On the other hand, we have already seen that the strategy in Thm. 18 is the local strategy that minimises q, and so any other fidelity estimation protocol that one might consider must incur a penalty that is even more acute.

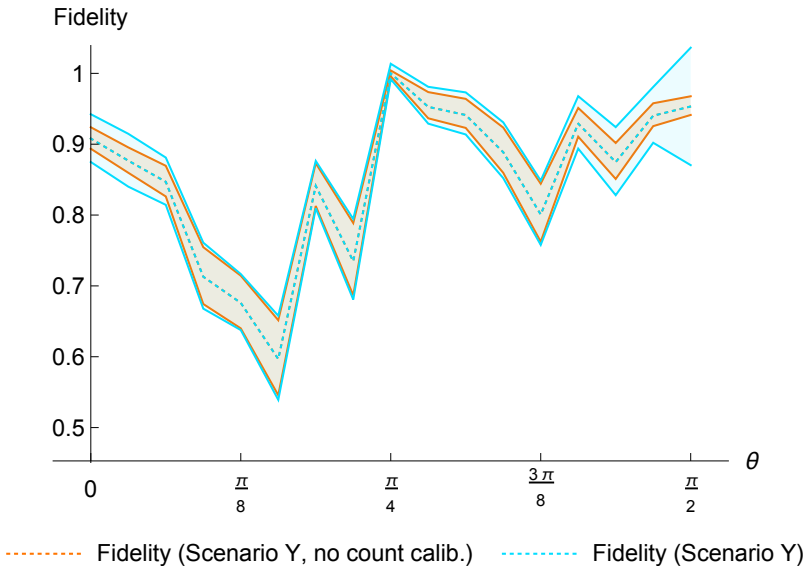

Figure 4.5: A comparison of the effect of miscalibrated counts. Both curves use the same data set; in the uncalibrated case, the number of counts for each setting are artificially scaled so that the weights μ_j are matched exactly, and confidence intervals are derived from Eq. 4.40. In the calibrated case, the confidence intervals are derived from Theorem 27.

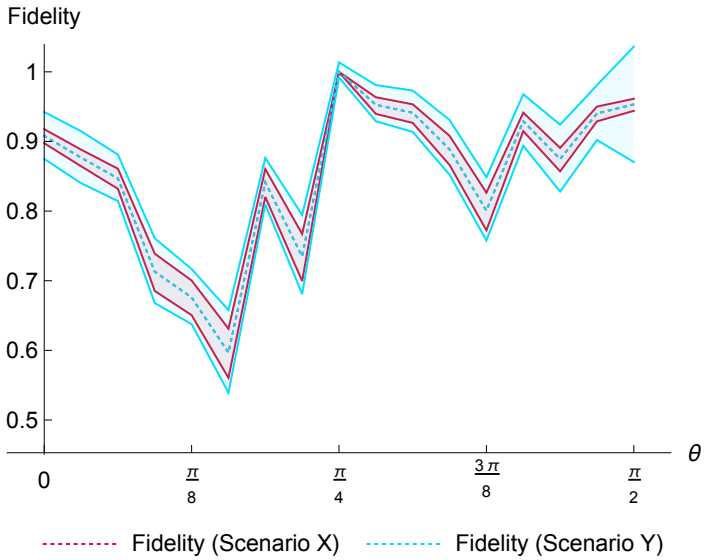

Figure 4.6: A comparison of the effect of implementing scenario Y over scenario X. The "scenario X" data set is the same as that for scenario Y, but where we have (falsely) made the assumption that each count is from a randomly selected measurement setting. The scenario X data confidence intervals do not include the effect of count miscalibration.

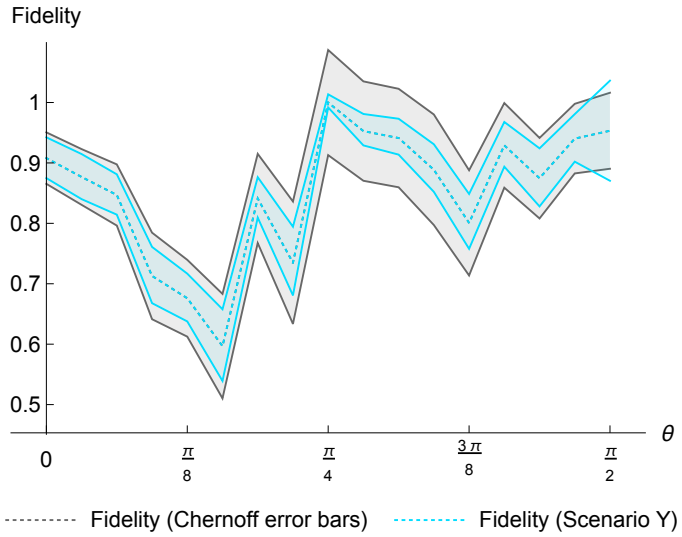

Figure 4.7: A comparison of the effect of choosing a looser concentration inequality, for the same data set (i.e. using the Chernoff bound in Eq. 4.46).

Regarding Point 3, it is clear that the problem of fluctuating counts is particularly troublesome for this data set. Despite each choice of θ being integrated for the same period of time, raw measurement counts vary by a factor of ~ 4 between the state $|11\rangle$ and the Bell state (see Fig. 4.4). Additionally, this fluctuation appears to be systematic rather than stochastic, and has a strong dependence on θ. This apparent dependence is problematic, as it only exacerbates the issue of growing errors bars for states close to a Bell state, given Point 1. It is unclear whether this variation could be tamed by better chip calibration, or whether it is a feature endemic in photonics experiments of this type.

The effect of count miscalibration is shown in Fig. 4.5. While the effect is small for $0 \leq \theta \leq \frac{\pi}{4}$, the effect becomes the dominant source of error for states close to either $|00\rangle$ or $|11\rangle$.

The effect of implementing scenario Y over scenario X is shown in Fig. 4.6. Broadly, switching scenarios would lead to $\sim 20\%$ smaller error bars, for almost all values of θ; however, implementing scenario X requires rapid switching of measurement bases which is likely to be infeasible given limitations of current technologies (or at least, would require a recalibration time of thermal phase shifters that would likely swamp any statistical advantage in integration time).

Finally, the effect of deriving error bars by a looser concentration inequality (in particular, Hoeffding's inequality in Eq. 4.46) is shown in Fig. 4.7. It is clear that this makes a significant difference on the size of the error bars, particularly for

high-fidelity states; the error bar for the Bell state, for example, is ~ 10 times larger when using Hoeffding's inequality than the tighter Chernoff-Hoeffding bound based on the relative entropy. It is worth noting that the prior art in fidelity estimation protocols [84, 213] both make use of this looser concentration inequality.

4.3.1 Comparison with other fidelity estimation protocols

To assess the quality and practicality of the fidelity estimation protocol derived above, we would like to compare it to previously established protocols (or bounds). To this end, we consider bounds on fidelity estimates from the precision-guaranteed tomography protocol in [213], and the direct fidelity estimation (DFE) protocol in [85].

Both of these protocols do not make use of the tailored measurement settings in Thm. 18, but instead rely on Pauli measurements to produce an estimate of the fidelity. To give as reasonable a comparison as possible for these bounds, we integrate for the same total integration time as the protocol derived here (but with integration time divided among 15 Pauli measurement settings, rather than the 4 settings in our protocol). If, after the same integration time, the total number of counts are fewer than those in Fig. 4.4, we artificially scale the total number of counts in favour of tomography and DFE, such that they agree. We make the generous assumption that both the fidelity estimation and tomography protocols were carried out completely perfectly (i.e. without error, miscalibration or loss).

To construct a point estimate for the fidelity in the tomographic case, we carry out the standard procedure for maximum likelihood tomography (see § 2.3.1).[1] We then directly compute the fidelity of the target state with the maximum likelihood state. In the direct fidelity estimation scenario, we carry out the protocol in § 2.1.4 with parameters given by Thm. 4. However, rather than choosing measurement settings at random for each trial, we fix blocks of measurement settings, with size normalised to the expectation of the target state for that particular Pauli observable.

Both of the error bar bounds are conservative, giving confidence interval guarantees for the worst case states and not relying on precise details of the data set. The precise details of the bounds we use are in § 2.1.4 (where we use the tighter bounds for "well-conditioned" states) and § 2.3.4, respectively. The resultant error bars are shown in Fig. 4.8. It is clear that, for this data set and for the practically implementable integration times achievable by this experiment, that the error bars derived from precision-guaranteed tomography and direct fidelity estimation in [85,

[1] As stated in § 2.3.4, the point estimate for precision-guaranteed tomography in [213] is technically not the maximum likelihood estimate, but the "extended-norm minimisation" estimate. However, as our focus will be on the error bars rather than the point estimate, we take the maximum likelihood estimate purely because it is easier to compute.

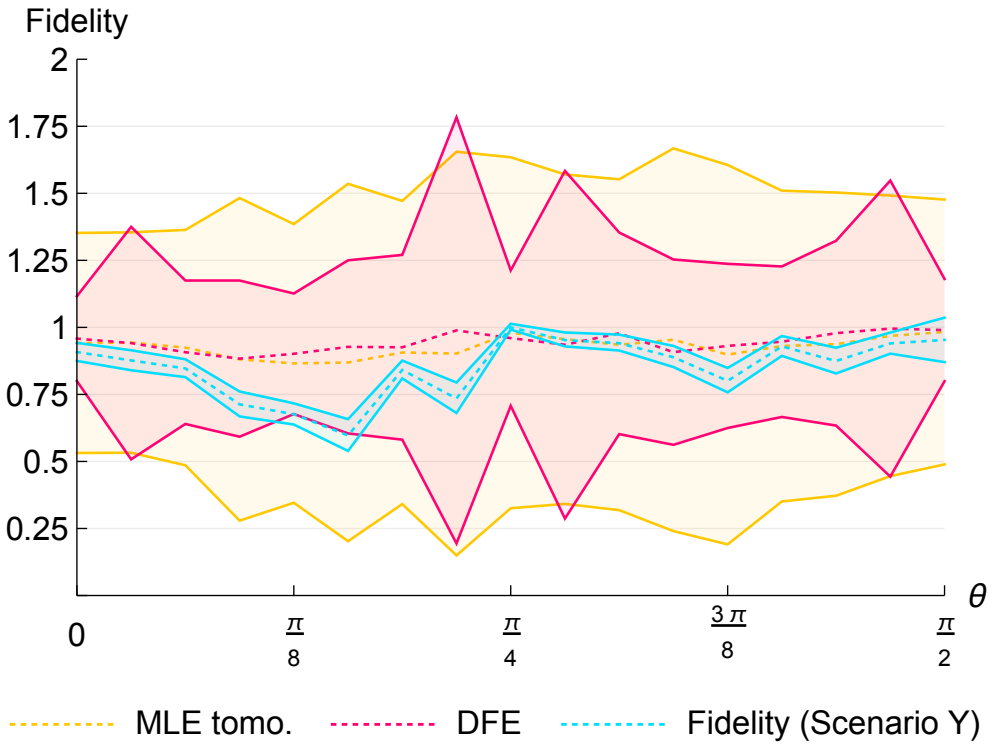

Figure 4.8: A comparison of the fidelity estimation protocol derived here, with the fidelity estimation protocol in [85] and the precision-guaranteed tomography protocol in [213].

213] are significantly larger than for the protocol derived here. It is also worth mentioning that the prior bounds are insensitive to the fact that the fidelity is bounded, $F \in [0,1]$, giving upper bounds that are significantly larger than the maximal fidelity across the entire data set. In general, it is arguable whether the prior tomography and fidelity estimation bounds have any predictive power at all for this data set, given that the error bars do not preclude both the exact target state with $F = 1$, and states close to the "worst-case" output, where the experiment just produces a maximally-mixed state, with $F = 0.25$.

We believe we have presented a protocol that is statistically rigorous given a plausible set of assumptions, and so the fact that our point estimate is significantly lower than in the tomography and DFE protocols, and strays close to the boundary of the DFE confidence interval around $\theta = \frac{\pi}{8}$, is problematic. A plausible explanation is to note that different measurement settings are used in each protocol, and it may well be the case that there is some systematic error in the way the chip applies the measurements in Thm. 18 that is not present for Pauli measurements, leading to lower point estimates in the fidelity. However if this is true, then there is some

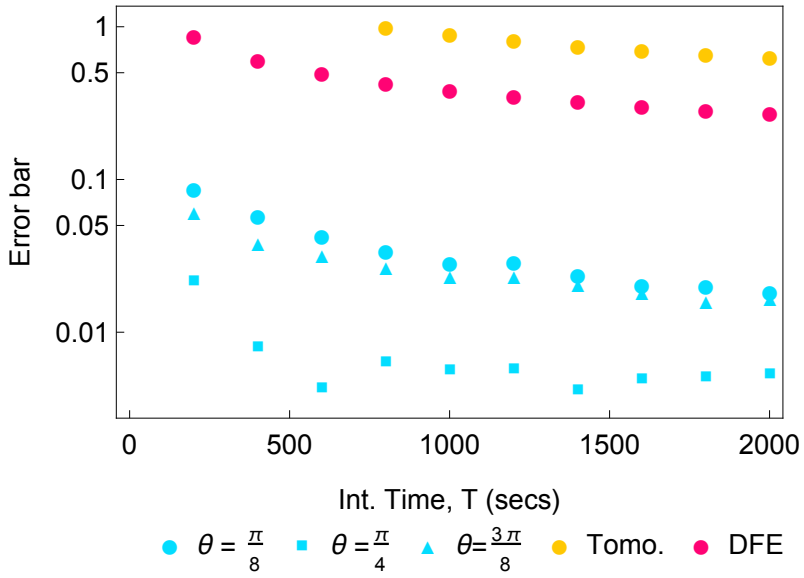

Figure 4.9: (i) Filled circles: a comparison of sizes of confidence interval in estimating the fidelity to $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$ for our protocol, tomography and direct fidelity estimation (DFE), for $\theta = \frac{\pi}{8}$, as a function of integration time. (ii) Cyan: a comparison of the sizes of confidence interval for our protocol, for three choices of θ.

physical flaw in the fabrication, calibration or operation of the chip that is entirely overlooked by the other protocols.

4.3.2 Performance with increasing integration time

It is also worthwhile considering how these error bars evolve when the integration time, and hence the total number of measurements performed, is increased. To this end, we carried out the protocol in Thm. 18 for varying integration times, ranging from 200 to 2000 seconds per choice of θ. To give a rough comparison with tomography and direct fidelity estimation as in Fig. 4.8, we take the total number of measurements for the original tomography data set integrated over 400 seconds and simply multiply by a scale factor to get an estimate for the total number of counts to be expected over the same range of integration times. The bound for precision-guaranteed tomography only depends on the total number of measurements; the direct fidelity estimation bound relies on both this, and the particular choice of θ.

The data is shown in Fig. 4.9, for our protocol taking $\theta \in \{\frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}\}$, for the DFE protocol taking $\theta = \frac{\pi}{8}$ and for the precision-guaranteed tomography protocol. We again take $\delta = 0.05$. At this δ and for an integration time $T \leq 600$ seconds, the implicit bounds from precision-guaranteed tomography have no solution. For all

choices of θ, we generically do not see an advantage in our protocol that grows with integration time. On the other hand, for each choice of θ, there is a significant constant-factor advantage in applying our protocol over the alternatives. For $\theta = \frac{\pi}{8}$, the error bars for our protocol are between $\sim 10\times$ and $\sim 15\times$ smaller than those for direct fidelity estimation, and between $\sim 30\times$ and $\sim 35\times$ smaller than those for precision-guaranteed tomography.

4.4 Overconfidence in photonic tomography

While there is a substantial body of work regarding confidence intervals in quantum state tomography (see § 2.3.4), it is very common, particularly in quantum photonics, to eschew these approaches in favour of techniques that are less statistically rigorous but simpler to implement. An overwhelming favourite is to apply a technique that we will refer to as "photonic estimation", the details of which we will describe shortly. This approach is typically flagged in the analysis by a phrase similar to "error bars calculated assuming Poissonian counting statistics" or "calculated by propagating errors that are assumed to be Poisson distributed" and is employed for a wide variety of estimated quantities, such as the fidelity, entanglement witnesses, and Hong-Ou-Mandel visibility [133, 207, 42, 226, 43, 228, 48, 198, 211, 197, 227, 229].

Suppose we have a setting where we measure multiple copies of some output state ρ, which we must tomographically reconstruct and then make use of this reconstruction to estimate some quantity Q. The protocol for deriving error bars using this method is to first construct a point estimate, then to use Monte Carlo methods to generate an artificial data set assuming that repeated experiments would give Poisson-distributed samples around the initial estimate. More specifically, the protocol is as follows:

Protocol Photonic estimation

for $i = 1$ to M **do**
  for $j = 1$ to $4^N - 1$ **do**
    From meas data construct point estimate $\tilde{P}_j$ of prob $P_j = \frac{1}{2}[\mathrm{tr}(\sigma_j \rho) + 1]$
    Take sample from artificial distribution $p_{ij} \sim \mathrm{Po}(\tilde{P}_j)$
    Store p_{ij} in $4^N - 1$- dim vector $\mathbf{p}_i$
  Calculate maximum likelihood state ρ_i consistent with the set $\mathbf{p}_i$
  Estimate the quantity Q given the state ρ_i, store in vec $\mathbf{Q}$
Output sample standard deviation of the elements of $\mathbf{Q}$.

The earliest result known to the author that uses this protocol in the context of continuous-variable tomography is due to Leonhardt et al. [139], and in the context

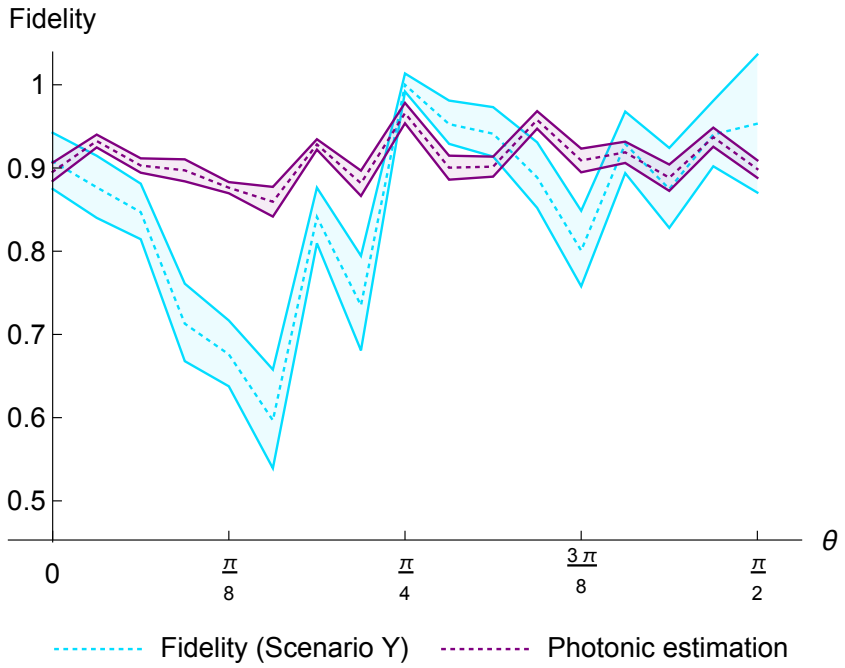

Figure 4.10: A comparison of the fidelity estimation protocol and the photonic estimation protocol in § 4.4.

of qubits by Hradil [117] and Altepeter, Kwiat and Jeffery [5].

However, this approach has a problem - it drastically overstates confidence when the point estimate $\tilde{P}_j$ is small. For example, suppose we know we have the single-qubit state $\rho = \alpha|0\rangle\langle 0| + (1-\alpha)|1\rangle\langle 1|$, for some unknown α. Now, suppose we measure Z on just a single copy, and get the outcome -1 (which occurs with probability $1-\alpha$). The maximum likelihood state consistent with the data is the state $\rho_{MLE} = |1\rangle\langle 1|$, and our maximum likelihood estimate of P_j is $\tilde{P}_j = \frac{1}{2}[\mathrm{tr}(\rho_{MLE} Z) + 1] = 0$. Hence the Monte Carlo samples are drawn from the distribution Po(0), a Poisson distribution with zero variance; and their sample standard deviation will be zero. The outcome is that, after taking only a single measurement and knowing essentially nothing about α, we are forced to conclude that the state was exactly $\rho = |1\rangle\langle 1|$, with absolute confidence.

The estimate $\tilde{P}_j$ can be small in two ways: either P_j isn't small, but we have not taken a sufficient number of measurements to guarantee that P_j and $\tilde{P}_j$ are close and have just been unlucky; or P_j itself is small. One could argue that the former is absolved when considering data sets of a reasonable size. However, a well-behaved confidence interval should, in principle, decrease monotonically in size with increasing data. This is not the case here - after a single measurement we have vanishing error bars, and if this artefact is to be corrected by taking more data, the error bars would have to experience an anomalous period of growth. The latter case

is more problematic; for most states of interest that we would like to verify, it is very often the case that we expect P_j to be small. Indeed, the DFE protocol in [85] explicitly uses this fact to derive an advantage over state tomography for useful quantum states (see the discussion on "well-conditioned" states in § 2.1.4).

In the case of output from the photonic chip, we can be more concrete about the magnitude of this problem. Given the tomographic data set used to construct Fig. 4.8, we can carry out the photonic estimation protocol, above. Point estimates are constructed by the same maximum likelihood technique as in Fig. 4.8, and errors bars are constructed from the photonic estimation protocol assuming $M = 100$ artificial Monte-Carlo states. The resultant comparison of the protocol derived in this chapter and photonic estimation is shown in Fig. 4.10. We expect for states of the form $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$ that most Pauli expectation values are close to zero, and that this problem is particularly apparent; and certainly, the data from the photonic estimation protocol in Fig. 4.10 is consistent with the hypothesis that the protocol systematically overstates the confidence in the point estimate. Across the entire data set, the error bar for photonic estimation is close to an order of magnitude smaller than for our fidelity estimation protocol. The fact that there is no overlap between the confidence intervals, particularly in the range $0 \leq \theta \leq \frac{\pi}{4}$, is enough to arouse suspicion. One could make the same counterargument as in § 2.5: different measurement settings are used in each protocol, and so it is plausible that there is some systematic error in the way the chip applies the measurements in Thm. 18 that is not present for Pauli measurements, leading to lower point estimates in the fidelity. Additionally, in this case there must also be some random error when these measurements are applied that is not present for Pauli measurements, leading to greater uncertainty in these estimates. However if this is true, we can draw the same conclusion: then there is some physical flaw in the fabrication, calibration or operation of the chip that is entirely overlooked by the photonic estimation protocol.

While this should not be considered an exhaustive discussion of the problems with photonic estimation, it should be enough to make the reader wary of any error bar derived in this manner. The problem is compounded by the fact that none of the following example papers: [133, 207, 42, 226, 43, 228, 48, 198, 211, 197, 227, 229], nor any paper that uses photonic estimation that is known to the author, either (i) present a statistical analysis of their technique for deriving error bars beyond a single phrase appealing to "Poisson statistics"; or (ii) provide a reference that discusses the analysis in more detail. This is particularly troubling when there is a substantial body of literature dedicated to deriving meaningful confidence intervals in quantum state tomography.

4.5 Outlook

The protocol we have derived and discussed above represents a clear step forward in the ability to estimate the fidelity of quantum states produced by a particular device or experiment. Broadly, the protocol is advantageous in the following respects: (i) given a fixed number of trials, it outputs significantly smaller error bars on its point estimate than its closest counterparts; (ii) it is constructed from non-Pauli measurements, and so may dig up coherent measurement errors that are not observed when carrying out "Pauli" tomography alone; and (iii) it requires a negligible amount of post-processing, in contrast with both maximum likelihood and Bayesian tomography. On the other hand, given that we have shown that the optimal verification protocols in Chapter 3 are also minimum-variance fidelity estimation protocols, we are hamstrung by the fact that we were not able to derive the optimal verification protocol for arbitrary pure states. It may well be the case that, even if deriving optimal fidelity estimation protocols is analytically taxing, that the framework in this chapter is sufficient to derive tight bounds on the performance of *any* candidate protocol. If this is the case, the most prudent course of action may be to explore ansatz protocols tailored to the state of interest.

We have also considered, in § 4.4, the most widely used procedure for deriving error bars on operational quantities in photonics. Based on both the statistical arguments and experimental data in § 4.4, it seems reasonable to conclude that this procedure significantly and quantifiably overstates the confidence in which it makes its point estimate. We hope that this evidence is sufficient for experimentalists in the field to revisit their use of this protocol, and to instead consider more statistically rigorous protocols (such as the one derived here, the direct fidelity estimation protocol in [85], or the tomographic estimators in [213, 26, 54]).

CHAPTER 5

QUANTUM ALGORITHMS FOR THE FINITE ELEMENT METHOD

The finite element method is used to approximately solve boundary value problems for differential equations. Qualitatively, the method discretises the parameter space and finds an approximate solution to the differential equation by solving a large system of linear equations. In this chapter, we investigate the extent to which the finite element method can be accelerated using an efficient quantum algorithm for solving linear equations. We consider the representative general question of approximately computing a linear functional of the solution to a boundary value problem, and compare the quantum algorithm's theoretical performance with that of a standard classical algorithm – the conjugate gradient method. We will find that the quantum algorithm can achieve a polynomial speedup, the extent of which grows with the dimension of the partial differential equation. On the other hand, this speed up is the most we can hope to achieve; we will see that no improvement of the quantum algorithm could lead to a super-polynomial speedup when the dimension is fixed and the solution satisfies certain smoothness properties.

5.1 Introduction

The development of a quantum algorithm for solving large systems of linear equations is an exciting recent advance in the field of quantum algorithmics. First introduced by Harrow, Hassidim and Lloyd [104], and later improved by other authors [6, 52, 232], the algorithm gives an exponential quantum speedup over classical algorithms for solving linear systems (see § 5.5 for a primer). However, the quantum linear equation (QLE) algorithm "solves" a system of equations $A\mathbf{x} = \mathbf{b}$ in an unusually quantum sense. The input $\mathbf{b}$ is provided as a quantum state $|b\rangle$, and the algorithm produces another state $|x\rangle$ corresponding to the desired output $\mathbf{x}$. Whether this is considered to be a reasonable definition of "solution" depends on the

intended application [1]. Still, linear equations are so ubiquitous in science and engineering that many applications of the QLE algorithm have been proposed, ranging from machine learning [142, 127, 128] to computing properties of electrical networks [225].

One area in which large systems of linear equations naturally occur is the finite element method (FEM) [10, 56, 34, 191]. The FEM is a technique for efficiently finding numerical approximations to the solutions of boundary value problems (BVPs) for partial differential equations, based on discretising the parameter space with a finite mesh. The FEM is a tempting target for acceleration by the QLE algorithm for several reasons:

1. The large systems of linear equations that occur in the FEM are produced algorithmically, rather than being given directly as input, which avoids efficiency issues associated with needing to access data via a quantum RAM [142, 1];

2. The FEM naturally leads to sparse systems of linear equations, which is usually a requirement for quantum speedup via the QLE algorithm;

3. The FEM gives an analytic expression for the condition number (a quantity characterising the "invertibility") of the matrix A, which can be folded in to the overall complexity;

4. The FEM has many important practical applications. These include structural mechanics, thermal physics and fluid dynamics [191]. Any quantum speedup for the FEM would thus represent a compelling application of quantum computers.

In this chapter we work through the details of applying the QLE algorithm to the general FEM, and compare the worst-case performance of the quantum algorithm with that of a ubiquitous classical algorithm.

5.1.1 Organisation and notation

Readers that wish to skip directly to the results comparing quantum and classical algorithms for the FEM will find them in § 5.7.

We begin, in § 5.2, by detailing the prior art regarding quantum algorithms for solving differential equations. We will then introduce the FEM by a simple example, before treating it more formally, in § 5.3. This framework will then be used to derive a complexity of a classical algorithm for the FEM in § 5.4. § 5.5 outlines the structure of the QLE algorithm, and § 5.6 goes through the details of applying the QLE algorithm to the FEM and determines its complexity. In § 5.8 we describe various limitations

on the quantum algorithm. We conclude in § 5.9 with some discussion and open problems.

We will need to deal with continuous functions, their discretised approximations as vectors, and their corresponding quantum states. Italics denote functions, boldface denotes vectors, and quantum states (usually normalised) are represented as kets. We often let $\Omega \subseteq \mathbb{R}^d$ denote an arbitrary convex set. For a function $f \in L^2(\Omega)$, $\|f\| := \left(\int_\Omega f(x)^2 dx\right)^{1/2}$ denotes the L^2 norm of f. For a vector $\mathbf{f}$, $\|\mathbf{f}\| := \left(\sum_i \mathbf{f}_i^2\right)^{1/2}$ denotes the ℓ_2 norm of $\mathbf{f}$.

If we call a solution to the PDE $u \in \mathbb{F}^d$, we will often use the term "spatial dimension" as shorthand for d, and as distinct from the dimension of the vector space used for a discretised approximation of the solution of a PDE, or the dimension of the Hilbert space acted on by a quantum algorithm. This is merely for convenience and should not be taken to mean that the only PDEs of interest are those in which the degrees of freedom are physical spatial dimensions.

5.2 Prior art

The application of the QLE algorithm to the FEM has previously been studied by Clader, Jacobs and Sprouse in [57]. In particular, they consider an electromagnetic scattering cross-section problem solved via the FEM, and argue that the quantum algorithm achieves an exponential speedup for this problem over the best classical algorithm known. In order to achieve this result, the authors of [57] propose ways to avoid issues with the QLE algorithm that can reduce or eliminate a quantum speedup. For example, they show that the important classical technique known as preconditioning, which reduces the condition number of the input matrix A, can be applied within the quantum algorithm. We discuss preconditioning further in § 5.3.

However, the analysis of [57] does not fully calculate and combine all contributions to the complexity of approximately solving the scattering cross-section problem. The classical and quantum algorithmic complexity is calculated in [57] in terms of two parameters: N (the size of the system of linear equations resulting from applying the FEM), and ϵ (the solution accuracy). The size of the system of equations is a parameter which can be chosen by the user in order to achieve a desired accuracy (i.e. N and ϵ are formally related). In [57] they are treated as independent parameters and hence the complexity analysis is left incomplete. If the scaling of N with ϵ is benign, the classical algorithm might not need to solve a large system of equations to achieve a given accuracy, so the quantum speedup could be reduced or even eliminated. This is the reason for the apparent contradiction between our results in § 5.7 and previous work [57].

Specifically, there are two potential sources of error in producing the solution: (i) the discretisation process which converts the differential equation to a system of linear equations; and (ii) any inaccuracies in numerically solving the system of equations itself and computing the desired function of the solution. The larger the system of equations produced, the smaller the error in (i) is. The QLE algorithm can work with an exponentially larger set of equations in a comparable time to the classical algorithm, so the error of type (i) can be made exponentially smaller. However, the scaling with accuracy of the QLE algorithm's extraction of a solution from the system of linear equations is substantially *worse* than the classical algorithm; the quantum algorithm has an exponentially larger error of type (ii). If only the error in (i) is considered, it may appear that the quantum algorithm yields an exponential advantage. However, the effect of both type (i) and type (ii) errors, when considered together, act in opposition and ultimately come close to cancelling each other out.

We should also note that a full resource analysis of the algorithm specified by Clader, Jacobs and Sprouse was carried out in [199].

An alternative approach to the approximate numerical solution of PDEs is the finite difference method (FDM). This method is also based on discretisation of the problem domain (in this instance, typically into a regular grid), but differs from the FEM in that it derives a linear system of equations from the PDE by approximating the partial derivatives in the original problem with finite differences.

The QLE algorithm can also be applied to the FDM. One example where this has been done, and described in detail, is work of Cao et al. [41], who gave a quantum algorithm for the Poisson equation in d dimensions. Their algorithm produces a quantum state corresponding to the solution to the equation in time $O(\max\{d, \log 1/\epsilon\} \log^3 1/\epsilon)$. Note that this scaling with ϵ is exponentially better than the best general results on Hamiltonian simulation known at the time; their algorithm used special properties of the Poisson equation to achieve an improved runtime. The best classical algorithms require time $\epsilon^{-\Omega(d)}$ as they solve a discretised version of the problem on a d-dimensional grid with cells of size $\epsilon \times \epsilon \times \cdots \times \epsilon$. However, the quantum algorithm of Cao et al. [41] shares the property of the FEM algorithms discussed here that, in order to extract some information from the quantum state produced, one finishes with a scaling with ϵ which is poly$(1/\epsilon)$. In the physically realistic setting of the dimension d being fixed and the accuracy ϵ being the parameter of interest, this is only a polynomial improvement.

Other related work has given quantum algorithms for solving large systems of sparse linear [21] or nonlinear [140] differential equations via Euler's method. In these cases the quantum algorithms can in principle achieve an exponential

improvement over classical computation for approximately computing properties of the solution to the system, if the system of equations is provided implicitly. Fleshing out this approach requires also specifying how the equations are produced and how the property of interest is computed. If the equations are generated by a discretisation procedure such as the FDM, similar qualitative conclusions to those we derive for the FEM seem likely to hold.

5.3 The finite element method

5.3.1 Warm-up: Poisson's equation

Rather than beginning by providing a formal introduction to the FEM, it is easiest to first motivate the procedure via an example. Imagine we would like to solve Poisson's equation on the interval $[0,1]$, in one dimension:

$$\frac{\partial^2 u}{\partial x^2} = f(x); \quad u(0) = \frac{\partial u}{\partial x}(1) = 0. \tag{5.1}$$

We will often use notation $'$ and $''$ to denote first and second derivatives, respectively. Here $f(x)$ is the input to the problem and we fix boundary conditions by specifying $u(0)$ and $u'(1)$. Given a sufficiently smooth "test function" $v \in L^2[0,1]$ such that $v(0) = 0$, one can multiply both sides by v and then integrate by parts:

$$\int_0^1 f(x)v(x)dx = \int_0^1 u''(x)v(x)dx = -\int_0^1 u'(x)v'(x)dx. \tag{5.2}$$

Assuming certain regularity properties of f, a function u which satisfies this equality for all test functions v will satisfy Poisson's equation. This is known as the *weak formulation* of Poisson's equation. The goal is to reduce this formulation to a problem that is tractable computationally. The approximation is to consider solutions and test functions that instead exist in some finite-dimensional subspace S of $L^2[0,1]$. Denote the approximate solution as $\tilde{u}$, such that $\tilde{u} \in S \subset L^2[0,1]$. Commonly S is taken to be the space of piecewise polynomial functions of some degree k; the choice of "pieces" for these functions is the origin of the finite element mesh.

A particularly simple choice of basis for this example is the space of piecewise linear functions on $[0,1]$, divided up into N intervals of size h. A basis for this space is the set of "tent" functions (see Fig. 5.1), defined as

$$\phi_i(x) = \begin{cases} \frac{1}{h}(x - x_{i-1}) & \text{if } x \in [x_{i-1}, x_i] \\ \frac{1}{h}(x_{i+1} - x) & \text{if } x \in [x_i, x_{i+1}] \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

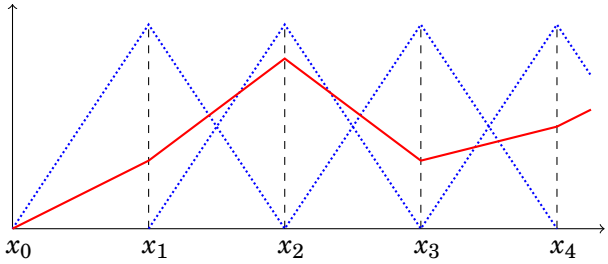

Figure 5.1: A basis set of "tent" functions (the blue, or dotted, lines), for piecewise linear functions defined on the line (an example of which is given by the red, or solid, line). Any piecewise linear function can be uniquely specified as a sum of scaled tents.

More generally, consider some choice of basis for the space S, denoted by $B = \{\phi_i\}$, such that $|B| = N$. We choose a basis such that $\phi_i(0) = \phi_i'(1) = 0$, so that every function in S satisfies the boundary conditions. Then $\tilde{u}$ can be expanded in this basis: $\tilde{u} = \sum_j U_j \phi_j$. The corresponding weak formulation of Poisson's equation is

$$-\sum_j U_j \int_0^1 \phi_j'(x) v'(x) dx = \int_0^1 f(x) v(x) dx. \tag{5.4}$$

For this condition to hold for all $v \in S$, it is sufficient for it to hold on all basis functions ϕ_i:

$$-\sum_j U_j \int_0^1 \phi_j'(x) \phi_i'(x) dx = \int_0^1 f(x) \phi_i(x) dx. \tag{5.5}$$

If we define N-dimensional vectors $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{f}}$ such that

$$\tilde{\mathbf{u}}_i = U_i, \quad \tilde{\mathbf{f}}_i = \int_0^1 f(x) \phi_i(x) dx \tag{5.6}$$

and an $N \times N$ matrix M such that

$$M_{ij} = \int_0^1 \phi_i'(x) \phi_j'(x) dx, \tag{5.7}$$

then the approximate solution to Poisson's equation can be determined by solving the linear system

$$M\tilde{\mathbf{u}} = \tilde{\mathbf{f}}. \tag{5.8}$$

5.3.2 The FEM for more complicated PDEs

This general procedure (expressing the PDE in the weak formulation, discretising by choosing a finite element mesh and basis functions and solving the resultant linear system of equations) can be extended to far more complicated PDEs, domains and boundary conditions. We will now give a brief pedagogical outline of the general

framework for each of these steps, for more complicated PDEs than the Poisson's equation example above.

We will be concerned with PDEs defined over some convex set $\Omega \subset \mathbb{R}^d$. Given that we are operating in more than one dimension, it will be useful to introduce a shorthand for derivatives with respect to multiple degrees of freedom. Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be a multi-index (i.e. a list of d non-negative integers). We'll use the shorthand $|\alpha| := \sum_i \alpha_i$, and denote mixed derivatives as $\partial^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \ldots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$. Thus each element α_j in the list α indicates the order of the derivative to be taken with respect to coordinate j.

The notion of a *weak derivative*, or of a *weak solution* as introduced in the above discussion on Poisson's equation, is designed to derive solutions with derivatives that need not exist, but nonetheless satisfy the equation in a particular sense. Broad classes of useful PDEs admit weak solutions that would not be reachable by other methods. The weak derivative generalises from the "integration by parts" procedure we used for the Poisson equation; the idea is to sidestep considering derivatives of the solution, by "transferring" them to derivatives of a test function. If we have some integrable function $g : \Omega \to \mathbb{R}$, then if there is an integrable function $h : \Omega \to \mathbb{R}$ such that

$$\int_\Omega h(x)v(x)\,dx = (-1)^{|\alpha|} \int_\Omega g(x)\partial^\alpha v(x)\,dx, \;\; \forall v \in C^\infty(\Omega), \tag{5.9}$$

then we say that h is a weak derivative of g. Note that this is exactly the same procedure we used previously; start from $h(x) = \partial^\alpha g(x)$, multiply by the test function $v(x)$, and then integrate. The prefactor $(-1)^{|\alpha|}$ is accrued from the minus signs picked up by integrating by parts $|\alpha|$ times.

Given the notion of weak derivatives, there is a natural norm that arises. We will use the notation $|\cdot|_m$ to denote the Sobolev (2-)seminorm

$$|v|_m := \left(\sum_{\alpha, |\alpha|=m} \|\partial^\alpha v\|^2 \right)^{1/2}. \tag{5.10}$$

That is, the seminorm acts like the L^2 norm over all partial derivatives of order m. The Sobolev m-norm is then simply defined by $\|v\|_m := \sum_{i=0}^m |v|_i$. Note that, for $m = 0$,

$$|v|_m = \|v\|_m = \|v\|. \tag{5.11}$$

Now, consider a linear BVP of the form

$$\mathcal{D}[u(x)] = \sum_\alpha c_\alpha(x)\partial^\alpha u(x) = f(x), \tag{5.12}$$

for some differential operator $\mathcal{D}$, some $x \in \mathbb{R}^d$, and some sufficiently smooth functions $c_\alpha(x)$. The weak formulation of this expression given some test function $v(x)$ is then

$$\int_\Omega f(x)v(x)\,dx = \int_\Omega \left[\sum_\alpha (-1)^{|\alpha|} \partial^\alpha (c_\alpha(x)v(x)) \right] u(x)\,dx := \int_\Omega \mathcal{D}'[v(x)]u(x)\,dx. \tag{5.13}$$

Denote the lefthand side as (f,v), the usual inner product for square-integrable functions, and the righthand side as $a(u,v)$. $a(u,v)$ is asymmetric, and so is not an inner product, but it is a bilinear form. If we let the space of test functions be $V = \{v \in L^2(\Omega) : a(v,v) < \infty\}$, then V is a subspace of the Hilbert space $L^2(\Omega)$, with inner product $(\cdot,\cdot)$ and norm $\|v\|_V = \sqrt{v,v}$. The weak formulation of the BVP is then succinctly written:

$$\text{Find } u \in V \text{ such that } a(u,v) = (f,v),\ \forall v \in V. \tag{5.14}$$

It is intuitive that if the BVP has a strong solution, that it must also satisfy Eq. 5.14. However if a strong solution does not exist or cannot be found, it is unclear whether Eq. 5.14 is a stringent enough formulation to lead to unique solutions. Luckily, the Lax-Milgram theorem (see [34], § 2.6, and [10], § 3.2) guarantees both the existence and uniqueness of solutions to the formulation in Eq. 5.14:

Theorem 28 (Lax-Milgram [135])**.** *Let V be a Hilbert space with bilinear form $a(\cdot,\cdot)$, such that a is both:*

1. ***Bounded:*** $|a(u,v)| \leq A\|u\|_V\|v\|_V$*, and*
2. ***Coercive:*** $|a(u,u)| \geq B\|u\|_V^2$*,*

for some constants A and B. Then for any f, there is a single unique solution to the problem

$$a(u,v) = (f,v). \tag{5.15}$$

The details of the domain Ω, the Hilbert space V, and the bilinear form $a(\cdot,\cdot)$ are governed by the BVP in question. Here we choose not to specify which PDE we wish to solve, as the details of this procedure for particular PDEs will not be very significant when making a general comparison of quantum and classical algorithms for the FEM. However, we will restrict to elliptic second-order PDEs throughout. The restriction to elliptic PDEs guarantees that the coercivity property of $a(\cdot,\cdot)$ is automatically satisfied, and so we do not need to worry about the existence and uniqueness of solutions to the PDE in the weak formulation, given the Lax-Milgram theorem. For further discussion on coercivity in PDEs, see [77]. Even with this restriction, the following analysis captures many examples of physical interest; for example, electrostatics, subsonic fluid dynamics and linear elasticity.

All that remains is to discretise the problem, to make it amenable to numerical methods. Given the form Eq. 5.14, its discretised form is remarkably simple to state; if we let S be a finite-dimensional subspace of V, then the problem becomes

$$\text{Find } \tilde{u} \in S \text{ such that } a(\tilde{u},v) = (f,v),\ \forall v \in S. \tag{5.16}$$

This is commonly known as the *Ritz-Galerkin* approximation. Not only is this approximation simple to state, but it is optimal, in some sense:

Theorem 29 (Cèa [44])**.** *Given that the conditions for the Lax-Milgram theorem are satisfied, and that the Ritz-Galerkin approximation to the solution u is denoted $\tilde{u}$, then*

$$\|u-\tilde{u}\|_V \leq \frac{A}{B}\|u-v\|_V,\ \forall v \in S. \tag{5.17}$$

Thus the Ritz-Galerkin solution $\tilde{u}$ is the closest approximation to the solution u in the finite subspace S (up to a constant).

The choice of the finite subspace S is the mathematical origin of the "finite elements". The uniform division of $[0,1]$ into equal intervals that we saw in the 1D case is replaced with a suitably regular division of the domain into a mesh, whose elements are usually polygons (for example, triangles) or polyhedra. An example of a mesh is shown in Fig. 5.2. The space S is replaced with the space of piecewise polynomials of degree k on the elements of the mesh, with a basis $\{\phi_i\}$ of polynomials supported only on adjacent mesh elements. Finally, the matrix M that we must invert, and is defined in Eq. 5.7 for the Poisson equation example, is modified such that $M_{ij} = a(\phi_i, \phi_j)$, where $a(u,v)$ is the bilinear form whose precise details depend on the PDE in question.

Preconditioning can be seen as replacing the matrix M with a matrix $M' = PM$ for some "preconditioner" P, and solving the new system of linear equations $M'\tilde{\mathbf{u}} = P\tilde{\mathbf{f}}$.

While preconditioners are commonly used as a subroutine in classical matrix inversion algorithms, one of the goals of the work by Clader, Jacobs and Sprouse [57] was to show that the sparse approximate inverse (SPAI) preconditioner can be used within the overall framework of the QLE algorithm. In the SPAI preconditioner, P is chosen such that $P \approx M^{-1}$ and also that P is sparse. The sparsity desired is a parameter of the algorithm; although one has no guarantees that either P or PM will be sparse while PM achieves a low condition number, in practice this is often the case. The structure of the SPAI is designed such that queries to entries of PM can be computed from queries to M with a modest overhead [57].

5.4 Solving the FEM with a classical algorithm

The goal of this section is derive a computational complexity that is representative of classical algorithms for solving BVPs via the FEM. However, as we will see in § 5.6, the quantum algorithm is hampered with respect to classical algorithms in that it does not allow the full solution u to a given BVP to be obtained, but does allow certain properties of u to be approximately computed. Thus in order to fairly compare classical and quantum algorithms for solving general BVPs in § 5.7, for both

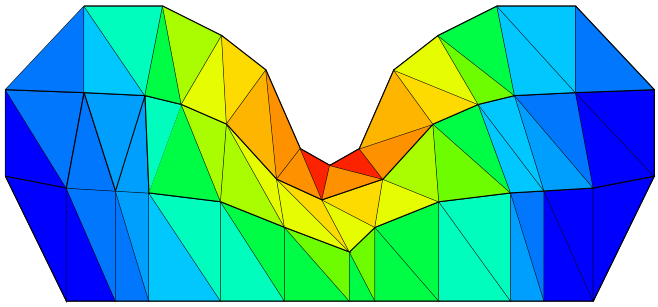

Figure 5.2: An example of a "mesh" – a discretisation of the domain over which the PDE is defined. Each polygon is a "finite element", with basis functions defined upon them. The shading of each polygon here represents the amplitude associated with the function supported on each finite element. As a physical example, the diagram could represent the material stress on a plate, induced by a deformation by a rod.

the classical and quantum case we consider the representative problem of computing a linear functional of u. That is, for some known function $r : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$, we seek to compute

$$\langle r, u \rangle := \int_\Omega r(\mathbf{x}) u(\mathbf{x}) d\mathbf{x}. \tag{5.18}$$

This is one of the simplest properties of u one could hope to access. In general, we do not have complete knowledge of u, but have some approximation $\tilde{u}$. Although there are many sensible norms with which one could measure the quality of this approximation, one natural choice is the L^2 norm $\|f\| := \left(\int_\Omega f(\mathbf{x})^2 d\mathbf{x})^2\right)^{1/2}$. Then

$$|\langle r, \tilde{u} \rangle - \langle r, u \rangle| = |\langle r, (\tilde{u} - u) \rangle| \leq \|r\| \|\tilde{u} - u\| \tag{5.19}$$

by the Cauchy-Schwarz inequality. Hence an accuracy of ϵ in L^2 norm in an approximation of u translates into an additive error of at most $\epsilon \|r\|$ in an approximation of $\langle r, u \rangle$. Therefore, approximating $\langle r, u \rangle$ up to accuracy $\epsilon \|r\|$ will be the prototypical problem considered throughout.

5.4.1 Approximation errors

If u is the exact solution to a BVP, henceforth let $\tilde{u}$ be the continuous, exact solution corresponding to the discretised problem in Eq. 5.8; $\tilde{u}$ is the solution that a perfect linear-system solver, given an arbitrarily long period of time, would find. In general, however, the linear-system solver is iterative and so will not truly reach $\tilde{u}$; so also let $\widetilde{\tilde{u}}$ be the continuous, approximate solution generated by the linear-system solver.

Crucially, one can show that $\tilde{u}$ can be made quite close to u by taking a sufficiently fine mesh. Indeed, consider a second order differential equation defined over a polygonal, d-dimensional domain (or equivalently, define d as the number of degrees of freedom in the PDE). Then, take an infinite, ordered family of

progressively finer meshes $\{\mathcal{M}_r\}_{r=1}^{\infty}$, constructed from a triangulation of the domain with simplices of dimension d. Let k be the total degree of the polynomials used as basis functions (so for example, $k = 1$ if we use the "tent" functions in Fig. 5.1). We assume throughout that both d and k are fixed. Given a parameter $m \in \mathbb{N}$, and provided that $d > 2(k - m)$, that all angles in the mesh are bounded below by some fixed value, and that the greatest edge length h in the mesh goes to zero, then the following bound is known ([56], Thm. 3.2.1):

$$|u - \tilde{u}|_m \leq C h^{k+1-m} |u|_{k+1}, \quad h \to 0, \tag{5.20}$$

assuming that weak derivatives of u of order m exist. Here C is a constant, independent of h (but not necessarily independent of d or the definition of the mesh). Setting $m = 0$ and given the equivalence of the Sobolev 0-seminorm, 0-norm and the L^2 norm in Eq. 5.11, we have $\|u - \tilde{u}\| \leq C h^{k+1} |u|_{k+1}$.

The overall level of inaccuracy in approximating u with $\widetilde{\tilde{u}}$ (and hence computing $\langle r, u \rangle$ from $\widetilde{\tilde{u}}$) can be bounded using the triangle inequality:

$$\|u - \widetilde{\tilde{u}}\| \leq \|u - \tilde{u}\| + \|\tilde{u} - \widetilde{\tilde{u}}\|. \tag{5.21}$$

To achieve a final error of $\epsilon \|r\|$ in computing $\langle r, u \rangle$ it is sufficient to achieve $\|u - \tilde{u}\| \leq \epsilon/2$, $\|\tilde{u} - \widetilde{\tilde{u}}\| \leq \epsilon/2$. Thus, by Eq. 5.20, we can take a mesh such that

$$h = O\left(\left(\frac{\epsilon}{|u|_{k+1}} \right)^{1/(k+1)} \right). \tag{5.22}$$

Observe that $|u|_{k+1}$ might be initially unknown. In the case of the simple instance of the FEM discussed in the previous section, we had $|u|_{k+1} = \|f\|$, so this bound could be explicitly calculated. However, it can be nontrivial to estimate this quantity for more complicated BVPs.

5.4.2 Classical complexity of the FEM

The overall complexity of solving a BVP via the FEM is governed by the dimensionality of the problem being solved, the choice of finite element basis, and the desired accuracy criteria. These feed into the complexity of solving the required system of linear equations.

As the matrix M to be inverted is a Gram matrix it is necessarily positive semidefinite. Also, the basis ϕ_i is almost universally chosen such that each basis vector only has support on a small number of finite elements, with the implication that M is sparse, i.e. has $s = O(1)$ nonzero entries in each row. The most common choice of algorithm for inversion of matrices of this type (large, sparse, symmetric and positive semidefinite) is the *conjugate gradient method* [205] (for discussion in

the context of the FEM, see [10], § 1.3). This method uses time $O(Ns\sqrt{\kappa}\log 1/\epsilon_{CG})$ to solve a system $M\tilde{\mathbf{u}} = \tilde{\mathbf{f}}$ of N linear equations, each containing at most s terms, with condition number $\kappa = \|M\|\|M^{-1}\|$, up to accuracy ϵ_{CG} in the "energy norm" $\|\mathbf{x}\|_M := \sqrt{\mathbf{x}^T M\mathbf{x}}$.

We now estimate the values of each of the parameters in this complexity, first calculating the required size N. Let $\mathcal{P}$ be a basis for the space of polynomials of total degree k in d variables. To construct a basis for the space of piecewise degree-k polynomials on the mesh, it is sufficient, for each finite element in the mesh, to include functions defined to be equal to a corresponding function in $\mathcal{P}$ on that finite element, and zero elsewhere. Then the total size of the basis is $N = O(h^{-d})$. Using Eq. 5.22, to achieve a final discretisation error of $\epsilon/2$ we can take

$$N = O\left(\left(\frac{|u|_{k+1}}{\epsilon}\right)^{\frac{d}{k+1}}\right). \tag{5.23}$$

We next determine the required accuracy ϵ_{CG}. Let a be the inner product defining M, such that $M_{ij} = a(\phi_i, \phi_j)$. This inner product induces the energy norm (on functions) $\|u\|_E := \sqrt{a(u,u)}$. Use of this norm makes it easy to interpret the error from the conjugate gradient method, as one can readily calculate that

$$\|\tilde{u} - \tilde{\tilde{u}}\|_E = \|\tilde{\mathbf{u}} - \tilde{\tilde{\mathbf{u}}}\|_M. \tag{5.24}$$

It follows from coercivity of a (see § 5.3) that $\|\tilde{u} - \tilde{\tilde{u}}\| \leq \sqrt{B}\|\tilde{u} - \tilde{\tilde{u}}\|_E$. To achieve $\|\tilde{u} - \tilde{\tilde{u}}\| \leq \epsilon/2$ it is therefore sufficient to take $\epsilon_{CG} = O(\epsilon)$.

The scaling of κ, the condition number of M, with the size and shape of the mesh is discussed extensively in [13] and [34, Chapter 9]. Assume that $d \geq 2$. Denote a particular triangulation of the domain Ω into finite elements as T, and let the expression $\mathrm{supp}(\phi_i) \cap T$ denote the set of finite elements where the basis function ϕ_i has support. Then we assume that there exists a universal constant C such that the basis functions ϕ_i satisfy

$$C^{-1}h^{d-2}\|v\|_{L^\infty(T)} \leq \sum_{\mathrm{supp}(\phi_i)\cap T \neq \emptyset} v_i^2 \leq Ch^{d-2}\|v\|_{L^\infty(T)} \tag{5.25}$$

for any function v such that $v = \sum_i v_i\phi_i$, and any triangulation T; this fixes the normalisation of the basis functions, and applies for all but the most irregular meshes. Then, for a wide range of relatively regular meshes, the largest eigenvalue $\lambda_{\max}(M) = O(1)$ and the smallest eigenvalue $\lambda_{\min}(M) = \Omega(N^{-2/d})$, so $\kappa = O(N^{2/d})$. Finally, we have $s = O(1)$ by our assumption about the supports of the basis elements ϕ_i. The overall complexity of the algorithm is thus

$$O\left(\left(\frac{|u|_{k+1}}{\epsilon}\right)^{\frac{d+1}{k+1}}\log\frac{1}{\epsilon}\right). \tag{5.26}$$

In many practical cases, however, we have seen that preconditioning is applied in order to reduce this scaling by improving the condition number. A number of different preconditioners are known; one frequently used example in the case of the FEM is the sparse approximate inverse (SPAI) preconditioner. Although there is no guarantee that this preconditioner can improve the condition number in the worst case, experimental results suggest that it can be very effective in practice [18, 19, 141, 182]. If the condition number were reduced to the best possible scaling $O(1)$, we would obtain a "best case" runtime of the classical algorithm which is

$$O\left(\left(\frac{|u|_{k+1}}{\epsilon}\right)^{\frac{d}{k+1}} \log\frac{1}{\epsilon}\right). \tag{5.27}$$

We remark that the preconditioned matrix M' may no longer be symmetric. The dependence of the conjugate gradient method on the condition number κ is quadratically worse for non-symmetric matrices, but as we have assumed that $\kappa = O(1)$ following preconditioning, this does not affect the complexity.

The best classical runtime following this approach is then found by optimising over allowed values of k. Observe that in either case, if $|u|_{k+1}$ and d are fixed, this complexity is bounded by a polynomial in $1/\epsilon$.

5.5 The HHL algorithm

We will now outline the current state of the art regarding quantum algorithms for solving linear systems of equations; it is these quantum algorithms that we will compare against classical iterative solvers, such as the conjugate gradient method. The first quantum algorithm for solving linear systems was published by Harrow, Hassidim and Lloyd [104] (and carries their initialism, "HHL"). The HHL algorithm produces the solution to the matrix equation $A\mathbf{x}' = \mathbf{b}$ in a particular, quantum sense. If we let A be an $N \times N$ matrix, then from the previous section we know that the conjugate gradient method (and most classical iterative methods) runs in time $O(N)$. But any algorithm, quantum or classical, that takes $\mathbf{b}$ as input and produces an approximation to $\mathbf{x}'$ as output needs time $O(N)$ just to read in, and write out, these vectors. Hence any fast quantum algorithm for inverting the matrix A will be swamped by the I/O time in this case, and any quantum advantage in N will vanish. Instead, the HHL algorithm assumes that we have oracular access to copies of a quantum state $|b\rangle = \mathbf{b}/\|\mathbf{b}\|$, which is somehow easy to produce (i.e. that it is produced in time $O(\mathrm{poly}\log(N))$). Additionally, the output is the matching quantum state over $\log(N)$ qubits, $|x\rangle = \mathbf{x}'/\|\mathbf{x}'\|$, rather than a vector of N bits. Given this setup, the runtime of the HHL algorithm is as follows:

Theorem 30 (Harrow, Hassidim and Lloyd [104], Ambainis [6])**.** *Let A be an $N \times N$ Hermitian matrix such that $\|A\|\|A^{-1}\| \leq \kappa$, and A has at most s nonzero entries in each row. Assume there is an algorithm $\mathcal{P}_A$ which, on input (r,i), outputs the location and value of the i^{th} nonzero entry in row r. Let $\mathbf{b}$ be an N-dimensional unit vector, and assume that there is an algorithm $\mathcal{P}_b$ which produces the corresponding state $|b\rangle$. Let*

$$\mathbf{x}' = A^{-1}\mathbf{b}, \quad |x\rangle = \frac{\mathbf{x}'}{\|\mathbf{x}'\|}. \tag{5.28}$$

Then there is a quantum algorithm which produces the state $|x\rangle$ up to accuracy ϵ in ℓ_2 norm, with bounded probability of failure, and makes the following number of calls to $\mathcal{P}_A$ and $\mathcal{P}_B$:

$$\mathcal{P}_A : O((s\kappa/\epsilon)\,\mathrm{poly}(\log(s\kappa/\epsilon))); \tag{5.29}$$

$$\mathcal{P}_B : O(s\kappa\,\mathrm{poly}(\log(s\kappa/\epsilon))). \tag{5.30}$$

The runtime is then

$$O((s\kappa/\epsilon)\,\mathrm{poly}(\log(s\kappa/\epsilon))\log(N)). \tag{5.31}$$

We will now give a brief flavour of the steps in the algorithm. We should point out that each step in the sketch below (producing $|b\rangle$, applying phase estimation and postselecting on the state of an ancilla) only produce approximations to the intermediate states we will describe; hence the analysis is significantly more involved than the quick summary here.

Let the initial input to the algorithm be the state $|b\rangle = \sum_i b_i|i\rangle$, where the amplitudes b_i are the (suitably normalised) entries in $\mathbf{b}$. Alternatively, let the eigenvalues and eigenvectors of A be labelled by λ_j and $|u_j\rangle$, respectively. Then we can rewrite $|b\rangle$ in this eigenbasis: $|b\rangle = \sum_j \beta_j|u_j\rangle$. Then, after using Hamiltonian simulation techniques to apply e^{iAt} and applying phase estimation, the output is $\sim \sum_j \beta_j|u_j\rangle|\lambda_j\rangle$. Our next goal is to apply the (non-unitary) map $|\lambda_j\rangle \to C\lambda_j^{-1}|\lambda_j\rangle$. We append an ancilla qubit, and perform a rotation on it conditioned on the register $|\lambda_j\rangle$, yielding $\sim \sum_j \beta_j|u_j\rangle|\lambda_j\rangle\left(\sqrt{1-C^2/\lambda_j^2}|0\rangle + C/\lambda_j|1\rangle\right)$. We then measure the ancilla in the computational basis and postselect on the measurement outcome -1; after this step, the output is $\sim \sum_j \beta_j\lambda_j^{-1}|u_j\rangle|\lambda_j\rangle|1\rangle$. All that is left is to discard the ancilla qubit, yielding a state $\sim \sum_j \beta_j\lambda_j^{-1}|u_j\rangle = |x\rangle$, as desired.

The dominant source of error in this algorithm comes from the phase estimation step, which runs in time $O(\kappa/\epsilon)$. The contribution by Ambainis [6] was to use variable-time amplitude amplification to reduce the dependence on κ in the overall algorithm from quadratic to linear. The authors of [104] are also able to place tight restrictions on any improvements to this algorithm. Specifically, if the dependence on κ is reduced to polylogarithmic then one can show BQP = PSPACE, and if the dependence on ϵ *for*

the task of outputting an expectation value over $|x\rangle$, rather than the state $|x\rangle$ itself is reduced to polylogarithmic in $1/\epsilon$, then BQP = PP. Both of these equalities are considered very unlikely. On the other hand, they also show that the problem of matrix inversion is BQP-complete, and so we do not expect any classical algorithm to match the polylogarithmic dependence on N of the HHL algorithm.

The original HHL algorithm has since been superseded by an improved algorithm by Childs, Kothari and Somma [52]:

Theorem 31 (Childs, Kothari and Somma [52])**.** *Let A be an $N \times N$ Hermitian matrix such that $\|A\|\|A^{-1}\| \leq \kappa$, and A has at most s nonzero entries in each row. Assume there is an algorithm $\mathcal{P}_A$ which, on input (r,i), outputs the location and value of the i'th nonzero entry in row r. Let $\mathbf{b}$ be an N-dimensional unit vector, and assume that there is an algorithm $\mathcal{P}_b$ which produces the corresponding state $|b\rangle$. Let*

$$\mathbf{x}' = A^{-1}\mathbf{b}, \quad |x\rangle = \frac{\mathbf{x}'}{\|\mathbf{x}'\|}. \tag{5.32}$$

Then there is a quantum algorithm which produces the state $|x\rangle$ up to accuracy ϵ in ℓ_2 norm, with bounded probability of failure, and makes

$$O(s\kappa \operatorname{poly}(\log(s\kappa/\epsilon))) \tag{5.33}$$

uses of $\mathcal{P}_A$ and $\mathcal{P}_b$. The runtime is the same up to a $\operatorname{poly}(\log N)$ *factor.*

In particular, for the task of producing the quantum state $|x\rangle$, this algorithm is exponentially faster in $1/\epsilon$ than the original HHL algorithm. The improvement over the prior algorithm is to note that the dependence on $1/\epsilon$ is wholly derived from phase estimation, and to sidestep its use for a much more direct method. Rather than phase estimation, the algorithm in [52] directly applies the matrix A^{-1} to the input state $|b\rangle$ by decomposing A^{-1} into a summation of efficiently implementable unitaries. They then give a procedure for applying this sum directly, that only requires $O(\operatorname{poly}\log(1/\epsilon))$ calls to $\mathcal{P}_A$ and $\mathcal{P}_B$.

There have been recent advances in algorithms for this problem, in particular results that apply even when the matrix is non-sparse [232]. Here, the algorithm retains the same dependence on κ and ϵ, but the dependence on the sparsity is replaced with a linear dependence on a quantity that is upper bounded by the Frobenius norm of A, $\|A\|_F$. In certain instances (for example, when A is dense but low rank), the sparsity could be $\Omega(N)$ and the Frobenius norm $O(1)$, which would imply an exponential advantage over the algorithm in [52]. However, the linear systems arising in the finite element method are $O(1)$ sparse, by design; as such, we will rely on the algorithm of Childs, Kothari and Somma [52] at the expense of a constant factor in runtime, in the worst case.

5.6 Solving the FEM with a quantum algorithm

The key step towards solving the FEM more quickly using a quantum computer is to replace the classical algorithm for solving the corresponding system of linear equations with the quantum algorithm in Theorem 31. We will also need to approximate the Euclidean norm of the solution, $\|\mathbf{x}'\|$. The most efficient approach to achieve this appears to be based on the original HHL algorithm. The number of uses of $\mathcal{P}_A$ required to estimate $\|\mathbf{x}'\|$ up to accuracy $\epsilon\|\mathbf{x}'\|$ can be shown to be

$$O((s\kappa^2/\epsilon)\,\mathrm{poly}\log(s\kappa/\epsilon)); \tag{5.34}$$

the number of uses of $\mathcal{P}_b$ required is $O(\kappa/\epsilon)$. Assuming that $\mathcal{P}_A$ and $\mathcal{P}_b$ can each be implemented in time $\mathrm{poly}(\log N)$, the runtime of the algorithm is

$$O((s\kappa^2/\epsilon)\,\mathrm{poly}\log(Ns\kappa/\epsilon)). \tag{5.35}$$

As we were unable to find statements of these bounds in the literature, we sketch the argument behind them in Appendix C.1.

Here we will apply these results to the linear system $M\tilde{\mathbf{u}} = \tilde{\mathbf{f}}$. We see from the above bounds that the complexity of the overall quantum algorithm for solving the FEM is determined by the following parameters:

1. The complexities of the algorithms $\mathcal{P}_M$ and $\mathcal{P}_{\tilde{f}}$, which respectively determine elements of M and (approximately) produce $|\tilde{f}\rangle$;

2. The condition number κ and sparsity s of the matrix M;

3. The complexity of determining some quantity of interest given a state which approximates $|\tilde{u}\rangle$.

These quantities will depend in turn on the desired accuracy of the output. We now investigate each of them.

Note that most of the algorithms we use will have some arbitrarily small, but non-zero, probability of failure. We assume throughout that failure probabilities have been made sufficiently low that they can be disregarded.

5.6.1 Preparing the input

The purpose of this section is to discuss the time to prepare the input state $|\tilde{f}\rangle$ (i.e. the complexity of the subroutine $\mathcal{P}_{\tilde{f}}$ required for the QLE algorithm). To achieve an efficient algorithm overall, we would like to be able to prepare $|\tilde{f}\rangle$ in time $\mathrm{poly}(\log N)$. Rather than rely on a quantum RAM to provide $|\tilde{f}\rangle$, we instead refer to

a scheme introduced by Zalka [236], and independently rediscovered by both Grover and Rudolph [98] and Kaye and Mosca [125].

The scheme can be used to produce a real quantum state $|\psi\rangle$ of n qubits in time polynomial in n, given the ability to compute the weights

$$W_x := \sum_{y \in \{0,1\}^{n-k}} |\langle xy|\psi\rangle|^2 \tag{5.36}$$

for arbitrary $k = 1,\ldots,n$ and arbitrary $x \in \{0,1\}^k$ in time poly(n), as well as the ability to determine the sign of $\langle x|\psi\rangle$ for arbitrary x in time poly(n).

To approximately produce $|\psi\rangle$ up to a high level of accuracy (e.g. $O(2^{-n})$) in time polynomial in n, it is actually sufficient to be able to approximately compute each weight W_x up to accuracy ϵ in time $O(\log 1/\epsilon)$, for arbitrary ϵ. We sketch the argument as follows. The algorithm of [236, 98, 125] is designed to produce a state $|\psi'\rangle$ with non-negative amplitudes in the computational basis, such that $\langle x|\psi'\rangle = |\langle x|\psi\rangle| = \sqrt{W_x}$ for all $x \in \{0,1\}^n$, and then flips the signs of amplitudes as required. To produce $|\psi'\rangle$ the algorithm expresses W_x, for each $x \in \{0,1\}^n$, as a telescoping product

$$W_x = W_{x_1} \times \frac{W_{x_1x_2}}{W_{x_1}} \times \frac{W_{x_1x_2x_3}}{W_{x_1x_2}} \times \cdots \times \frac{W_x}{W_{x_1\ldots x_{n-1}}}, \tag{5.37}$$

computes each fraction in turn (in superposition), and uses this to set $\langle x|\psi'\rangle$. If the goal is to produce $|\psi\rangle$ up to accuracy ϵ in ℓ_2 norm, from the inequality $(|\langle x|\psi\rangle| - |\langle x|\psi'\rangle|)^2 \leq |\langle x|\psi\rangle^2 - \langle x|\psi'\rangle^2|$ it is sufficient to approximate each weight W_x, $x \in \{0,1\}^n$, up to additive error $O(\epsilon^2/2^n)$. So the product can be truncated at the point i where the weight $W_{x_1\ldots x_i} = O(\epsilon^2/2^n)$, because any subsequent multiplications can only decrease W_x, and weights below this size can be ignored.

If the algorithm does not compute weights W_x, W_y in some fraction W_x/W_y exactly, but instead computes approximations $\widetilde{W_x}$ and $\widetilde{W_y}$ such that $|\widetilde{W_x} - W_x| \leq \gamma W_x$ and $|\widetilde{W_y} - W_y| \leq \gamma W_y$ for some γ, then $|\widetilde{W_x}/\widetilde{W_y} - W_x/W_y| = O(\gamma W_x/W_y)$. As we have assumed that $W_x = \Omega(\epsilon^2/2^n)$ for all k-bit strings x for which we compute W_x ($1 \leq k \leq n$), it is sufficient to approximate each weight W_x up to additive accuracy $O(\epsilon^2/(n2^n))$ for each fraction to be accurate up to a multiplicative error of $O(\epsilon^2/(n2^n))$ and hence the overall product of weights to be accurate up to an additive error $O(\epsilon^2/2^n)$. From the assumption about the complexity of the algorithm for approximately computing W_x, we can achieve this level of accuracy in poly($n, \log 1/\epsilon$) time.

In the case of the FEM, the weights W_x correspond to quantities of the form

$$S(a,b) := \sum_{i=a}^{b} \left(\int_\Omega \phi_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \right)^2, \tag{5.38}$$

where $\mathbf{x} \in \mathbb{R}^d$ and a and b are integers. Expressions of this form can be computed (either exactly or approximately) for many functions f of interest. For example,

consider the 1-dimensional setting discussed in § 5.3. If f is a polynomial, then the integral can be easily calculated, and corresponds to a polynomial in x_{i-1}, x_i and x_{i+1}. If the finite elements are regularly spaced, so $x_i = ih$ for some h, the entire sum $S(a,b)$ is a polynomial in a and b which can be explicitly calculated for any a and b.

For a choice of polynomial basis of degree k (i.e. where the $(k+1)^{th}$ derivative $\phi_i^{(k+1)} = 0$), then from Darboux's formula one has that

$$\int_\Omega \phi_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \sum_{j=1}^{k} (-1)^j \phi_i^{(j)}(\mathbf{x}) \underbrace{\int \cdots \int}_{j+1 \text{ times}} f(\mathbf{x}) d\mathbf{x}. \tag{5.39}$$

So, once the basis is specified, computing individual amplitudes is only as difficult as integrating the function $f(\mathbf{x})$. However, the state-production algorithm requires the computation of weights which depend on up to $N = 2^n$ squared amplitudes. To obtain an efficient algorithm, it is therefore necessary to find a more concise expression for these sums.

As discussed above, this can be achieved when f is a polynomial and the finite element mesh is suitably regular. This includes some physically interesting cases; even a constant function f can be of interest. An efficiently computable expression for $S(a,b)$ can also be obtained when f is only supported on a few basis elements. However, it appears challenging to compute this quantity efficiently for more general functions f. Indeed, see § 5.8.3 for an argument that this should not be achievable in general.

For simplicity in the subsequent bounds, we henceforth assume that the state $|\tilde{f}\rangle$ can be produced perfectly in time $O(\text{poly}\log N)$.

5.6.2 Solving the system of linear equations

Let M be the matrix defined by $M_{ij} = a(\phi_i, \phi_j)$. Recall from Theorem 31 that the quantum algorithm assumes that it has access to an algorithm $\mathcal{P}_M$ which, on input (r,i), outputs the location and value of the i'th nonzero entry in row r (or "not found" if there are fewer than i nonzero entries). If the finite element mesh is suitably regular, $\mathcal{P}_M$ is easy to implement. For instance, consider the set of n piecewise linear functions on $[0,1]$ defined in § 5.3. Then M is a tridiagonal matrix whose diagonal elements are equal to $2/h$, and whose off-diagonal elements are equal to $-1/h$. Hence for $r > 1$,

$$\mathcal{P}_M(r,i) = \begin{cases} (r-1, -1/h) & \text{if } i = 1 \\ (r, 2/h) & \text{if } i = 2 \\ (r+1, -1/h) & \text{if } i = 3, \\ \text{not found} & \text{otherwise.} \end{cases} \tag{5.40}$$

More generally, it will be possible to implement $\mathcal{P}_M$ efficiently if there is an efficient procedure for mapping an index to a finite element, and for listing the neighbouring elements for a given element. This will be the case, for example, when the finite element mesh is a regular triangulation of a polygon. For discussion on automated mesh generation and indexing schemes, see both [101] and [10, § 5.1].

When solving the system of linear equations, inaccuracies in the prepared state $|\tilde{f}\rangle$ will translate into inaccuracies in the output state $|\tilde{u}\rangle$. Let $|\widetilde{\tilde{f}}\rangle$ be the approximate state that was actually prepared. Then the state produced after applying the QLE algorithm is (approximately)

$$\frac{M^{-1}|\widetilde{\tilde{f}}\rangle}{\|M^{-1}|\widetilde{\tilde{f}}\rangle\|}. \tag{5.41}$$

If $|\tilde{f}\rangle$ is prepared up to accuracy ϵ in ℓ_2 norm, then the inaccuracy of the output state in ℓ_2 norm is

$$\left\| \frac{M^{-1}|\tilde{f}\rangle}{\|M^{-1}|\tilde{f}\rangle\|} - \frac{M^{-1}|\widetilde{\tilde{f}}\rangle}{\|M^{-1}|\widetilde{\tilde{f}}\rangle\|} \right\|. \tag{5.42}$$

Writing $|\widetilde{\tilde{f}}\rangle = |\tilde{f}\rangle + |\epsilon\rangle$ for some vector $|\epsilon\rangle$ such that $\||\epsilon\rangle\| = \epsilon$, this quantity is equal to

$$\begin{aligned}
&\left\| \frac{M^{-1}|\tilde{f}\rangle(\|M^{-1}|\widetilde{\tilde{f}}\rangle\| - \|M^{-1}|\tilde{f}\rangle\|)}{\|M^{-1}|\tilde{f}\rangle\|\|M^{-1}|\widetilde{\tilde{f}}\rangle\|} - \frac{M^{-1}|\epsilon\rangle}{\|M^{-1}|\widetilde{\tilde{f}}\rangle\|} \right\| \\
&\leq \frac{|\|M^{-1}|\widetilde{\tilde{f}}\rangle\| - \|M^{-1}|\tilde{f}\rangle\||}{\|M^{-1}|\widetilde{\tilde{f}}\rangle\|} + \frac{\|M^{-1}|\epsilon\rangle\|}{\|M^{-1}|\widetilde{\tilde{f}}\rangle\|} && (5.43) \\
&\leq 2\frac{\|M^{-1}|\epsilon\rangle\|}{\|M^{-1}|\widetilde{\tilde{f}}\rangle\|} && (5.44) \\
&\leq 2\epsilon\kappa, && (5.45)
\end{aligned}$$

by the triangle inequality, the reverse triangle inequality, and the definition of the condition number κ. We therefore see that, if preconditioning is not applied to the matrix M to reduce κ, it is necessary for $|\tilde{f}\rangle$ to be prepared up to accuracy $O(N^{-2/d}\epsilon)$. (Note that this is not an issue if we can produce $|\tilde{f}\rangle$ exactly, as for some examples discussed in the previous section.)

If preconditioning is used, we no longer need to prepare the initial state $|\tilde{f}\rangle$, but a state proportional to $P|\tilde{f}\rangle$. Note that preparing the input $P\tilde{\mathbf{f}}$ in the classical case requires only multiplication of a vector by a sparse matrix, which is computationally cheap compared to matrix inversion. As such, it is typically neglected when considering the classical computational complexity. However, the situation is more complicated in the quantum setting.

The most straightforward way to prepare $P|\tilde{f}\rangle$ is to construct $|\tilde{f}\rangle$ and then attempt to apply the (non-unitary) operation P. There are several known approaches which can be used to achieve this probabilistically. One elegant example is a simple special

case of the "Chebyshev" approach of [52, §4]. This work uses a quantum walk to apply n^{th} order Chebyshev polynomials $T_n(P)$ in an arbitrary s-sparse Hermitian matrix P. (If P is not Hermitian, a standard trick [104] can be used to express it as a submatrix of a Hermitian matrix.) As the first Chebyshev polynomial T_1 is simply $T_1(x) = x$, this allows P itself to be implemented. If the subroutine of [52] succeeds when applied to a state $|\psi\rangle$, then $|\psi\rangle$ is (exactly) mapped to $P|\psi\rangle/\|P|\psi\rangle\|$. The success probability is at least

$$\frac{\|P|\psi\rangle\|^2}{s^2\|P\|_{\max}^2} \geq \frac{1}{\kappa(P)^2 s^2}, \tag{5.46}$$

where $\|P\|_{\max} = \max_{i,j}|P_{ij}|$ and we use $\|P\|_{\max} \leq \|P\|$. Using amplitude amplification, the failure probability can be made at most δ, for arbitrarily small $\delta > 0$, with $O(1/(\kappa(P)s))$ repetitions. Each repetition requires time polylogarithmic in N, $\kappa(P)$ and s.

Combining all these considerations, we see that if preconditioning is used, the complexity of the quantum algorithm will depend on a number of different parameters, each of which may be hard to estimate in advance. These are: the condition number of PM; the complexity of computing entries of P; the sparsity of P; and the condition number of P. Here, in order to give a "best case" comparison of the preconditioned quantum algorithm with the classical algorithm, we make the optimistic assumption that preconditioning is optimal (i.e. $\kappa(PM) = O(1)$), and that taking all of these additional sources of complexity into account multiplies the runtime by only a poly(log(N)) factor.

5.6.3 Measuring the output

By running the QLE algorithm, we obtain an output state $|\widetilde{u}\rangle$ which approximates the normalised state

$$|\tilde{u}\rangle = \frac{\sum_i \tilde{\mathbf{u}}_i |i\rangle}{\sqrt{\sum_i \tilde{\mathbf{u}}_i^2}}, \tag{5.47}$$

where we associate each basis state $|i\rangle$ with the basis function ϕ_i. Given copies of $|\tilde{u}\rangle$, we can carry out measurements to extract information about u. One example is the prototypical problem we consider here, approximating the L^2 inner product $\langle u, r\rangle$ between u and a fixed function r. This can be achieved by approximately computing the inner product $\langle\tilde{u}|r\rangle$ between $|\tilde{u}\rangle$ and the state $|r\rangle$ defined by

$$|r\rangle = \frac{1}{(\sum_i \langle\phi_i, r\rangle^2)^{1/2}} \sum_i \langle\phi_i, r\rangle |i\rangle \tag{5.48}$$

for some function r; then $\langle\tilde{u}|r\rangle$ is the L^2 inner product between $\tilde{u}$ and r, up to an overall scaling factor. $|r\rangle$ can be produced using techniques described in the previous

section. Some interesting cases are particularly simple: for example, taking r to be uniform on a region, $\langle \tilde{u}|r\rangle$ gives the average of $\tilde{u}$ over that region.

This inner product can be estimated using a procedure known as the Hadamard test [4], a subroutine whose output is a ± 1-valued random variable with expectation $\langle \tilde{u}|r\rangle$. By applying amplitude estimation [31] to approximately compute this expectation, $\langle \tilde{u}|r\rangle$ can be estimated up to accuracy ϵ with $O(1/\epsilon)$ uses of algorithms to produce the states $|\tilde{u}\rangle$ and $|r\rangle$. A related approach was used by Clader, Jacobs and Sprouse [57] to compute an electromagnetic scattering cross-section, which corresponds to a quantity of the form $|\langle \tilde{u}|r\rangle|^2$. This can be approximately computed using the swap test [38], a subroutine which, given two states $|\psi\rangle$, $|\psi'\rangle$, outputs "same" with probability $\frac{1}{2} + \frac{1}{2}|\langle \psi|\psi'\rangle|^2$, and "different" otherwise.

We remark that more complicated properties of u seem to be more problematic to compute directly from the state $|\tilde{u}\rangle$, due to the non-orthogonality of the basis $\{\phi_i\}$. For example, one common use of the QLE algorithm is to determine similarity of solutions to sets of linear equations by using the swap or Hadamard tests to compare them [104, 6]. Consider two states $|a\rangle$, $|b\rangle$ corresponding to functions $a = \sum_i \mathbf{a}_i \phi_i$, $b = \sum_i \mathbf{b}_i \phi_i$. Then

$$\langle a|b\rangle \propto \sum_i \mathbf{a}_i \mathbf{b}_i \tag{5.49}$$

while a sensible measure of similarity of the functions a and b is the inner product

$$\int_\Omega a(\mathbf{x}) b(\mathbf{x}) d\mathbf{x} = \sum_{i,j} \mathbf{a}_i \mathbf{b}_j \int_\Omega \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) d\mathbf{x}. \tag{5.50}$$

One would hope for this to be approximately proportional to $\sum_i \mathbf{a}_i \mathbf{b}_i$. However, although ϕ_i and ϕ_j do not have overlapping support for most pairs $i \neq j$, there are still enough such pairs where this overlap is nonzero that the integral can sometimes be a poor approximation.

5.6.4 Overall complexity

The total complexity of the quantum algorithm for solving an FEM problem is found by combining the complexities of all of the above pieces.

Assume that we would like to compute $R := \int_\Omega r(\mathbf{x}) u(\mathbf{x}) d\mathbf{x}$ for some $r : \Omega \to \mathbb{R}$ up to additive error $\epsilon \|r\|$. Write $\alpha = (\sum_i \langle \phi_i, r\rangle^2)^{1/2}$. The quantum algorithm will perform the following steps by applying the QLE algorithm to the system of linear equations $M\tilde{\mathbf{u}} = \tilde{\mathbf{f}}$:

1. Estimate $\|\tilde{\mathbf{u}}\|$ up to an additive term ϵ_N. Let $\widetilde{N}$ be the estimate.

2. Use the QLE algorithm to produce copies of $|\widetilde{\tilde{u}}\rangle$, an approximation to $|\tilde{u}\rangle$. Use these to estimate $\langle r|\widetilde{\tilde{u}}\rangle$ up to an additive term ϵ_{out}. Let $\widetilde{R}$ be the estimate.

3. Output $\alpha\widetilde{N}\widetilde{R}$ as an estimate of R.

We can bound the overall error as follows. Let ϵ_L be the inaccuracy, in ℓ_2 norm, in solving the system of linear equations in step (2), i.e. $\epsilon_L = \||\widetilde{\tilde{u}}\rangle - |\tilde{u}\rangle\|$. This encompasses any error in producing the initial state $|\tilde{f}\rangle$, as well as inaccuracy arising from the QLE algorithm itself (although recall that we have in fact assumed that we can produce $|\tilde{f}\rangle$ perfectly). Then

$$\begin{aligned}\widetilde{R} &= \langle r|\widetilde{\tilde{u}}\rangle + \epsilon_{\text{out}} \\ &= \langle r|\tilde{u}\rangle + \langle r|(|\widetilde{\tilde{u}}\rangle - |\tilde{u}\rangle) + \epsilon_{\text{out}} \\ &= \langle r|\tilde{u}\rangle + \epsilon'_L + \epsilon_{\text{out}} \end{aligned} \tag{5.51}$$

for some ϵ'_L, where $|\epsilon'_L| \leq \epsilon_L$ by Cauchy-Schwarz. So

$$\widetilde{R} = \frac{\sum_i \tilde{\mathbf{u}}_i \langle \phi_i, r\rangle}{\|\tilde{\mathbf{u}}\| \left(\sum_i \langle \phi_i, r\rangle^2\right)^{1/2}} + \epsilon'_L + \epsilon_{\text{out}} = \frac{\langle \tilde{u}, r\rangle}{\alpha\|\tilde{\mathbf{u}}\|} + \epsilon'_L + \epsilon_{\text{out}} \tag{5.52}$$

using the definition of $|r\rangle$ from (5.48) and of $|\tilde{u}\rangle$ as a normalised version of $\tilde{\mathbf{u}}$. Writing $\widetilde{N} = \|\tilde{\mathbf{u}}\| + \epsilon_N$, we have

$$\alpha\widetilde{N}\widetilde{R} = \langle \tilde{u}, r\rangle \left(1 + \frac{\epsilon_N}{\|\tilde{\mathbf{u}}\|}\right) + \alpha(\|\tilde{\mathbf{u}}\| + \epsilon_N)(\epsilon'_L + \epsilon_{\text{out}}). \tag{5.53}$$

The analysis of the remaining term $\langle \tilde{u}, r\rangle$ is now similar to the classical setting:

$$\langle \tilde{u}, r\rangle = \langle u, r\rangle + \langle \tilde{u} - u, r\rangle = \langle u, r\rangle + \epsilon'_D, \tag{5.54}$$

where

$$|\epsilon'_D| \leq \|r\|\|\tilde{u} - u\| =: \|r\|\epsilon_D \tag{5.55}$$

by Cauchy-Schwarz. Combining all the terms together, we have

$$\alpha\widetilde{N}\widetilde{R} - R = \epsilon'_D \left(1 + \frac{\epsilon_N}{\|\tilde{\mathbf{u}}\|}\right) + \frac{\langle u, r\rangle \epsilon_N}{\|\tilde{\mathbf{u}}\|} + \alpha(\|\tilde{\mathbf{u}}\| + \epsilon_N)(\epsilon'_L + \epsilon_{\text{out}}). \tag{5.56}$$

We finally use $\langle u, r\rangle \leq \|u\|\|r\|$ to obtain

$$\alpha\widetilde{N}\widetilde{R} - R = \epsilon'_D \left(1 + \frac{\epsilon_N}{\|\tilde{\mathbf{u}}\|}\right) + \frac{\|u\|\|r\|\epsilon'_N}{\|\tilde{\mathbf{u}}\|} + \alpha(\|\tilde{\mathbf{u}}\| + \epsilon_N)(\epsilon'_L + \epsilon_{\text{out}}) \tag{5.57}$$

for some ϵ'_N such that $|\epsilon'_N| \leq |\epsilon_N|$. To achieve overall accuracy $\epsilon\|r\|$ it is sufficient for each term in this expression to be upper-bounded by $\epsilon\|r\|/3$, which follows from

$$\epsilon_D = O(\epsilon), \tag{5.58}$$

$$\epsilon_N = O(\min\{\|\tilde{\mathbf{u}}\|, \epsilon\|\tilde{\mathbf{u}}\|/\|u\|\}), \tag{5.59}$$

$$\epsilon_L, \epsilon_{\text{out}} = O(\epsilon\|r\|/(\alpha\|\tilde{\mathbf{u}}\|)). \tag{5.60}$$

Assume for simplicity in the final bound that $\epsilon \leq \|u\|$; then the second condition becomes $\epsilon_N = O(\epsilon\|\tilde{\mathbf{u}}\|/\|u\|)$. We now calculate the complexity of achieving these accuracies.

Using the discretisation error bound in Eq. 5.20, we have $\epsilon_D \leq C h^{k+1}|u|_{k+1}$ for some universal constant C. We can therefore take

$$h = O\left(\left(\frac{\epsilon}{|u|_{k+1}}\right)^{\frac{1}{k+1}}\right). \tag{5.61}$$

As in the classical case, this choice of h corresponds to solving a system of

$$N = O\left(\left(\frac{|u|_{k+1}}{\epsilon}\right)^{\frac{d}{k+1}}\right) \tag{5.62}$$

linear equations. For any $\delta > 0$, as discussed in § 5.6 and Appendix C.1, $\|\tilde{\mathbf{u}}\|$ can be approximated up to accuracy $\delta\|\tilde{\mathbf{u}}\|$ in time $O((s\kappa^2/\delta)\,\mathrm{poly}\log(Ns\kappa/\delta))$ using the HHL algorithm, recalling that s and κ are the sparsity and condition number of M, respectively. Inserting $\delta = \epsilon/\|u\|$ and the bound on N, this part requires time

$$O\left(\frac{s\kappa^2\|u\|}{\epsilon}\,\mathrm{poly}(\log(s\kappa\|u\||u|_{k+1}/\epsilon))\right) \tag{5.63}$$

We can also put an upper bound on ϵ_L and ϵ_{out} by upper-bounding $\alpha/\|r\|$ and $\|\tilde{\mathbf{u}}\|$. In the former case, it holds that

$$\frac{\alpha}{\|r\|} = O(h\sqrt{s}); \tag{5.64}$$

we prove this technical claim in Appendix C.2. In the latter case,

$$\|\tilde{\mathbf{u}}\| = O(\|\tilde{\mathbf{u}}\|_M/h) = O(\|\tilde{u}\|_E/h) = O(\|u\|_1/h). \tag{5.65}$$

The first two equalities follow from $\lambda_{\min}(M) = \Omega(N^{-2/d}) = \Omega(h^2)$ [13, 34] and the equivalence between $\|\tilde{\mathbf{u}}\|_M$ and $\|\tilde{u}\|_E$ (see Eq. 5.24). The third follows from $\|\tilde{u}\|_E = O(\|\tilde{u}\|_1)$, where $\|\cdot\|_1$ is the Sobolev 1-norm, which is a consequence of the inner product $a(\cdot,\cdot)$ defining the energy norm corresponding to an underlying elliptic PDE [34], and the bound in Eq. 5.20. Combining these bounds, the requirement on ϵ_L and ϵ_{out} can be rewritten as

$$\epsilon_L, \epsilon_{\text{out}} = O\left(\frac{\epsilon}{\sqrt{s}\|u\|_1}\right). \tag{5.66}$$

To achieve accuracy ϵ_{out} using the Hadamard test and amplitude estimation requires $O(1/\epsilon_{\text{out}})$ uses of the QLE algorithm, each of which runs in time $O(s\kappa\,\mathrm{poly}(\log(Ns\kappa/\epsilon_L)))$ by Theorem 31. Inserting the bounds on ϵ_L, ϵ_{out} gives a complexity for part (2) which is

$$O\left(\frac{\sqrt{s}\kappa\|u\|_1}{\epsilon}\,\mathrm{poly}(\log(s\kappa\|u\|_1|u|_{k+1}/\epsilon))\right) \tag{5.67}$$

Combining these bounds, we obtain an overall runtime of

$$O\left(\frac{s\kappa^2\|u\| + \sqrt{s}\kappa\|u\|_1}{\epsilon}\,\mathrm{poly}(\log(s\kappa\|u\|_1|u|_{k+1}/\epsilon))\right). \tag{5.68}$$

In fixed spatial dimension, $s = O(1)$, and if preconditioning is not used, $\kappa = O(N^{2/d}) = O((|u|_{k+1}/\epsilon)^{2/(k+1)})$. Inserting these values, we obtain a bound of

$$\widetilde{O}\left(\frac{\|u\||u|_{k+1}^{\frac{4}{k+1}}}{\epsilon^{\frac{k+5}{k+1}}} + \frac{\|u\|_1|u|_{k+1}^{\frac{2}{k+1}}}{\epsilon^{\frac{k+3}{k+1}}}\right), \tag{5.69}$$

where the $\widetilde{O}$ notation hides polylogarithmic factors. On the other hand, if we assume that optimal preconditioning has been applied, so κ reduces to $O(1)$, we would obtain an overall bound of just

$$\widetilde{O}\left(\frac{\|u\|_1}{\epsilon}\right). \tag{5.70}$$

These runtimes should be compared with the corresponding runtimes of the classical algorithm, in the cases of no preconditioning and optimal preconditioning, respectively:

$$\widetilde{O}\left(\left(\frac{|u|_{k+1}}{\epsilon}\right)^{\frac{d+1}{k+1}}\right) \text{ and } \widetilde{O}\left(\left(\frac{|u|_{k+1}}{\epsilon}\right)^{\frac{d}{k+1}}\right). \tag{5.71}$$

5.7 Comparing quantum and classical algorithms for the FEM

Examples of the bounds we have derived are listed in Table 5.1, which includes the effect of preconditioning on the runtime of the algorithms (see § 5.3 for the definition of preconditioning). Note that in general it is difficult to rigorously analyse the performance of preconditioners. We therefore choose to highlight two extreme possibilities: no preconditioning at all is applied, or maximally successful preconditioning is used. The true performance of an algorithm using preconditioning will fall somewhere between the two cases.

First, note that if preconditioning is used, the dependence on the smoothness of the solution in the quantum algorithm's runtime is significantly milder than for the classical algorithm, being only polylogarithmic. The runtime of both the classical and quantum algorithms depends on the Sobolev ℓ-seminorm and Sobolev ℓ-norm of the solution to the BVP, for some ℓ; which as we have seen, measure the size of the ℓ^{th} derivatives of the solution. Assuming that preconditioning has been optimally used within the QLE algorithm, the quantum algorithm's runtime is dependent only on the Sobolev 1-norm (up to polylogarithmic terms). However, the classical algorithm's runtime depends on the Sobolev ℓ-seminorm, for some $\ell \geq 2$. Therefore,

Algorithm	No preconditioning	Optimal preconditioning
Classical	$\widetilde{O}\left((\lvert u\rvert_2/\epsilon)^{(d+1)/2}\right)$	$\widetilde{O}\left((\lvert u\rvert_2/\epsilon)^{d/2}\right)$
Quantum	$\widetilde{O}\left(\lVert u\rVert \lvert u\rvert_2^2/\epsilon^3 + \lVert u\rVert_1 \lvert u\rvert_2/\epsilon^2\right)$	$\widetilde{O}(\lVert u\rVert_1/\epsilon)$

Table 5.1: Complexity comparison of the algorithms studied in this chapter. Quantities listed are the worst-case time complexities of approximating a linear functional of the solution to a d-dimensional BVP up to accuracy ϵ, using the FEM with linear basis functions (see § 5.6 for bounds when using higher-degree polynomials). $\lVert u\rVert$, $\lvert u\rvert_\ell$ and $\lVert u\rVert_\ell$ are the L^2 norm, Sobolev ℓ-seminorm and Sobolev ℓ-norm of the solution respectively, defined in § 5.3. The $\widetilde{O}$ notation hides polylogarithmic factors.

for problems with solutions whose higher-order derivatives are large, the quantum advantage could be substantial. Even if preconditioning is not used, for large enough d the dependence is polynomially better.

Perhaps more importantly, observe that in the term dependent on ϵ in the algorithms' runtimes, the quantum algorithm's runtime no longer depends on the dimension d. This holds whether or not preconditioning is used. Thus the quantum algorithm will achieve a large speedup when ϵ is small and d is large. (Note that we cannot quite call this an exponential quantum speedup with respect to d: the runtime of the quantum algorithm also depends on a term C in Eq. 5.20 which is constant for a fixed dimension and family of meshes, but has unbounded dependence on d.) One example application where this is the case is any dynamical problem involving n bodies, which implies solving a PDE defined over a configuration space of dimension $2n$. In this case the quantum algorithm has a polynomial advantage, with the exponent equal to the number of bodies. Also, there may be a significant advantage for problems in mathematical finance; for example, pricing multi-asset options requires solving the Black-Scholes equation over a domain with dimension given by the number of assets [122]. Here, the quantum advantage is proportional to the number of assets in the option. This is discussed further in § 5.9.

As a demonstrative example of the speedup (or otherwise) expected for the quantum algorithm in fixed dimensions, consider the case of solving a BVP in four dimensions (three spatial and one temporal, say), using piecewise linear basis functions. Then, the classical runtimes both without preconditioning and with optimal preconditioning are

$$\widetilde{O}\left(\left(\frac{\lvert u\rvert_2}{\epsilon}\right)^{\frac{5}{2}}\right) \text{ and } \widetilde{O}\left(\left(\frac{\lvert u\rvert_2}{\epsilon}\right)^{2}\right), \tag{5.72}$$

respectively. The analogous quantum runtimes are

$$\widetilde{O}\left(\frac{\|u\||u|_2^2}{\epsilon^3}+\frac{\|u\|_1|u|_2}{\epsilon^2}\right) \text{ and } \widetilde{O}\left(\frac{\|u\|_1}{\epsilon}\right). \tag{5.73}$$

In this case, lack of preconditioning leads to a quantum algorithm which might or might not outperform the classical algorithm, depending on the relative sizes of ϵ, $\|u\|$, $\|u\|_1$ and $|u|_2$. In the optimally preconditioned case, the quantum algorithm both scales better with accuracy and has a less stringent condition on the solution smoothness.

The results we have derived can be summarised as follows: we find that the QLE algorithm is indeed applicable to the general FEM, and can achieve substantial speedups over the classical algorithm. However, the quantum speedup obtained is only at most polynomial, if the spatial dimension is fixed and the solution satisfies certain smoothness properties. For example, the maximal advantage of the quantum algorithm for the typical physically relevant PDE defined over 3+1 dimensions (three spatial and one temporal, such that $d = 4$) is approximately quadratic. In small enough dimension, and if the solution is sufficiently smooth, the runtime of the quantum algorithm can actually be worse than the classical algorithm.

We also remind the reader that there is a subtle point here regarding the output of the quantum algorithm: the scaling with accuracy of the quantum algorithm is substantially better if we only wish to produce the quantum state corresponding to the solution to the FEM [52], rather than computing some property of the state by measuring it. However, in applications one will always eventually want to perform a measurement to extract information from the final output of the quantum algorithm. We therefore have considered it reasonable to compare the quantum and classical complexities of producing a (classical) answer to some given problem.

Our results pinpoint the regimes in which one can hope to achieve exponential quantum speedups for the FEM, and show that apparent speedups can disappear when one takes the effect of solution accuracy into account. Nevertheless, we believe that the fact that exponential speedups might still be obtained in some cases is encouraging, and an incentive to focus on problems with a possibility of a genuine exponential quantum speedup.

5.8 Quantum lower bounds

Finally, we will argue that the inability of the quantum algorithm to deliver exponential speedups (in some cases) is not a limitation of the algorithm itself, but rather that any quantum algorithm for the FEM will face similar constraints. We elucidate several barriers with which any quantum algorithm will have to contend.

First, we show that, informally, any algorithm which needs to distinguish between two states which are distance ϵ apart must have runtime $\Omega(1/\sqrt{\epsilon})$. Second, we argue that the "FEM solving subroutine" of any quantum algorithm can likely be replaced with an equivalent classical subroutine with at most a polynomial slowdown (in fixed spatial dimension and when the solution is smooth). Third, we show that there can be no more than a quadratic speedup if the input to the problem is arbitrary and accessed via queries to a black box or "oracle".

5.8.1 A general quantum lower bound

We observe from Theorem 31 and the discussion in § 5.6.3 that producing the quantum state $|x\rangle \propto A^{-1}|b\rangle$ for some well-conditioned, sparse matrix A can be achieved in time $\mathrm{poly}\log(1/\epsilon)$, while the apparently simpler task of approximating some natural properties of $|x\rangle$ uses time $O(1/\epsilon)$. It is therefore natural to suspect that the runtime of this component could be improved substantially, e.g. to $\mathrm{poly}\log(1/\epsilon)$. However, it was shown by Harrow, Hassidim and Lloyd [104] that the existence of a quantum algorithm with this scaling for approximating some very simple properties of $|x\rangle$ would imply the complexity-theoretic consequence BQP=PP, which is considered highly unlikely (implying, for example, that quantum computers could efficiently solve NP-complete problems).

As well as this complexity-theoretic argument, we now give an argument based on ideas from query complexity which lower bounds the runtime of any algorithm which approximates some function of the output of the QLE algorithm, without making use of the internal structure of the algorithm. This encompasses all the uses of QLE for the FEM discussed in § 5.6.3.

We adapt a standard technique of Bennett et al. [17]. Consider an algorithm which has access to a unitary subroutine $\mathcal{A}_\psi$, parametrised by an unknown state $|\psi\rangle$, such that $\mathcal{A}_\psi$ maps $|0\rangle$ to $|\psi\rangle$. The algorithm may also have access to the inverse subroutine $\mathcal{A}_\psi^{-1}$. The algorithm does not know anything about how $\mathcal{A}_\psi$ is implemented and uses it as a "black box". It aims to estimate some property of $|\psi\rangle$. In the context of the FEM, we think of $\mathcal{A}_\psi$ as the QLE algorithm, where $|\psi\rangle$ is the output state corresponding to the approximate solution of the desired BVP. We assume that the algorithm only makes use of the QLE subroutine for one instance, i.e. it uses $\mathcal{A}_\psi$ throughout, rather than $\mathcal{A}_{\psi'}$ for some $|\psi'\rangle \neq |\psi\rangle$; relaxing this assumption would only make the problem harder.

For notational simplicity, we also assume in the proof that the overall algorithm does not use $\mathcal{A}_\psi^{-1}$ and does not use any ancilla qubits. (These assumptions can easily be relaxed without changing the conclusions.) Further assume that the overall algorithm makes T uses of $\mathcal{A}_\psi$, interspersed with arbitrary unitary operators

$U_1,\ldots,U_{T+1}$. Let $|\phi\rangle$ be such that $\||\psi\rangle - |\phi\rangle\| \leq \epsilon$, and such that the output of the algorithm should be different when using $\mathcal{A}_\phi$ rather than $\mathcal{A}_\psi$. Finally let $|\eta\rangle_{\psi,t}$ be the state of the overall algorithm after t uses of $\mathcal{A}_\psi$. Then

$$\begin{aligned}
\||\eta\rangle_{\psi,T} - |\eta\rangle_{\phi,T}\| &= \|U_{T+1}\mathcal{A}_\psi U_T \ldots \mathcal{A}_\psi U_1|0\rangle - U_{T+1}\mathcal{A}_\phi U_T \ldots \mathcal{A}_\phi U_1|0\rangle\| \\
&= \|\mathcal{A}_\psi U_T \ldots \mathcal{A}_\psi U_1|0\rangle - \mathcal{A}_\phi U_T \ldots \mathcal{A}_\phi U_1|0\rangle\| \\
&\leq \|\mathcal{A}_\psi U_T \mathcal{A}_\psi U_{T-1}\mathcal{A}_\psi \ldots \mathcal{A}_\psi U_1|0\rangle - \mathcal{A}_\psi U_T \mathcal{A}_\phi U_{T-1}\mathcal{A}_\phi \ldots \mathcal{A}_\phi U_1|0\rangle\| \\
&+ \|\mathcal{A}_\psi U_T \mathcal{A}_\phi U_{T-1}\mathcal{A}_\phi \ldots \mathcal{A}_\phi U_1|0\rangle - \mathcal{A}_\phi U_T \ldots \mathcal{A}_\phi U_1|0\rangle\| \\
&\leq \|\mathcal{A}_\psi U_{T-1}\mathcal{A}_\psi \ldots \mathcal{A}_\psi U_1|0\rangle - \mathcal{A}_\phi U_{T-1}\mathcal{A}_\phi \ldots \mathcal{A}_\phi U_1|0\rangle\| + \|\mathcal{A}_\psi - \mathcal{A}_\phi\|.
\end{aligned} \tag{5.74}$$

The first inequality is the triangle inequality, while the second uses the fact that unitaries do not change the Euclidean distance. As the algorithm does not use any information about the internal structure of $\mathcal{A}_\psi$, $\mathcal{A}_\phi$, we are free to assume that $\mathcal{A}_\psi = |\psi\rangle\langle 0| + |\phi'\rangle\langle 1| + \sum_{i\geq 2}|\zeta_i\rangle\langle i|$, $\mathcal{A}_\phi = |\phi\rangle\langle 0| + |\psi'\rangle\langle 1| + \sum_{i\geq 2}|\zeta_i\rangle\langle i|$. Here $|\phi'\rangle$ and $|\psi'\rangle$ are states orthonormal to $|\psi\rangle$ and $|\phi\rangle$, respectively, within the subspace spanned by $|\psi\rangle$ and $|\phi\rangle$, and $|\zeta_i\rangle$ are arbitrary states which are orthonormal to both of these states and each other. Explicitly, we can take

$$|\phi'\rangle = \frac{|\phi\rangle - \langle\psi|\phi\rangle|\psi\rangle}{\sqrt{1-|\langle\psi|\phi\rangle|^2}},\ |\psi'\rangle = \frac{|\psi\rangle - \langle\phi|\psi\rangle|\phi\rangle}{\sqrt{1-|\langle\psi|\phi\rangle|^2}}. \tag{5.75}$$

Then

$$\|\mathcal{A}_\psi - \mathcal{A}_\phi\| = \|(|\psi\rangle - |\phi\rangle)\langle 0| + (|\phi'\rangle - |\psi'\rangle)\langle 1|\|. \tag{5.76}$$

Writing $|\delta\rangle := |\psi\rangle - |\phi\rangle$, $|\delta'\rangle := |\phi'\rangle - |\psi'\rangle$ and upper-bounding the operator norm by the Frobenius norm, we have

$$\begin{aligned}
\|\mathcal{A}_\psi - \mathcal{A}_\phi\| &\leq \sqrt{\mathrm{tr}(\mathcal{A}_\psi^\dagger - \mathcal{A}_\phi^\dagger)(\mathcal{A}_\psi - \mathcal{A}_\phi)} \\
&= \sqrt{\langle\delta|\delta\rangle + \langle\delta'|\delta'\rangle} \\
&= \sqrt{2}\||\psi\rangle - |\phi\rangle\|,
\end{aligned} \tag{5.77}$$

where we use the fact (which can easily be seen by direct calculation) that $\||\phi'\rangle - |\psi'\rangle\| = \||\psi\rangle - |\phi\rangle\|$. Hence

$$\|\mathcal{A}_\psi - \mathcal{A}_\phi\| \leq \sqrt{2}\epsilon \tag{5.78}$$

and in turn, by induction,

$$\||\eta\rangle_{\psi,T} - |\eta\rangle_{\phi,T}\| \leq T\sqrt{2}\epsilon. \tag{5.79}$$

As the algorithm is supposed to output something different if it is given $\mathcal{A}_\phi$ rather than $\mathcal{A}_\psi$, assuming that it succeeds, the final measurement made distinguishes

between the two states $|\eta\rangle_{\psi,T}$ and $|\eta\rangle_{\phi,T}$. The optimal worst-case probability p of distinguishing these states is given by the trace distance between them [112], so

$$p = \frac{1}{2} + \frac{1}{4}\|\eta_{\psi,T} - \eta_{\phi,T}\|_1 \leq \frac{1}{2} + \frac{1}{2}\||\eta\rangle_{\psi,T} - |\eta\rangle_{\phi,T}\| \leq \frac{1}{2} + \frac{T\epsilon}{\sqrt{2}}. \tag{5.80}$$

Therefore, in order for the algorithm to succeed with probability (say) 2/3, it must use $\mathcal{A}_\psi$ at least $\Omega(1/\epsilon)$ times. As a simple example of how this bound can be applied, consider an algorithm which attempts to distinguish between these two cases: a) the output from the QLE subroutine is a particular state $|\psi_0\rangle$; b) the output from the QLE subroutine is some state $|\phi\rangle$ such that the overlap $|\langle\phi|\psi_0\rangle|^2 = 1-\epsilon$. Then $\||\phi\rangle - |\psi_0\rangle\| = O(\sqrt{\epsilon})$, so any algorithm distinguishing between these two cases by using QLE as a black box must use it $\Omega(1/\sqrt{\epsilon})$ times.

This bound is tight for this particular problem, which can be solved by using the QLE subroutine $O(1/\sqrt{\epsilon})$ times within quantum amplitude estimation [31]. However, for other problems it may be possible to put stronger lower bounds on the complexity.

5.8.2 Replacing the QLE subroutine with a classical algorithm

The above lower bound shows, roughly speaking, that any algorithm which uses the QLE subroutine as a black box and attempts to determine up to accuracy ϵ some property of the output state must make $\Omega(1/\sqrt{\epsilon})$ uses of the subroutine. However, in some cases it can be of interest to approximate properties of the output state to quite low levels of accuracy.

For example, consider the problem of distinguishing between the following two cases: a) the solution to a BVP is periodic; b) the solution is far from periodic. As it is known that quantum algorithms can test periodicity of functions exponentially faster than classical algorithms can [45], one might hope to use QLE, together with the quantum periodicity tester, to solve this problem exponentially faster than any classical algorithm.

Also note that it is likely to be hard to prove that it is impossible to obtain a superpolynomial quantum speedup for solving BVPs, if we define "solving" a BVP as computing an arbitrary function of the solution to a BVP. For example, we could contrive a BVP where the solution is easy to write down, and can be interpreted as an integer; and could then ask the algorithm to output the prime factors of that integer. Proving that quantum computers could not outperform classical computers for this task would imply an efficient classical algorithm for integer factorisation.

Nevertheless we believe that, even given a quantum algorithm for solving problems of this form, any uses of the QLE algorithm as a subroutine could be replaced with a classical algorithm, with at most a polynomial slowdown if the spatial dimension is fixed and the solution is suitably smooth. This would imply

that any exponential quantum speedup in the overall algorithm is not due to the part of it that solves the FEM. Making this argument rigorous seems challenging for technical reasons related to regularity of meshes and comparing different norms to measure accuracy, so we do not attempt it here, instead merely sketching the ideas informally.

The argument proceeds as follows. Imagine we have an overall quantum algorithm which uses the QLE algorithm as a subroutine to solve T FEM instances in spatial dimension bounded by $d = O(1)$, such that the solution to each instance has all relevant Sobolev norms bounded by $O(1)$. Then each such instance can be approximately solved by a classical algorithm using a mesh of size $\text{poly}(1/\epsilon)$, for any desired accuracy ϵ. We replace each subroutine which applies the QLE algorithm to solve an instance of the FEM, using a mesh $\mathcal{M}$ to achieve accuracy ϵ, with the following procedure:

1. Classically solve the same FEM instance, using a mesh $\mathcal{M}'$ which achieves accuracy $\max\{\gamma/T, \epsilon\}$ for some universal constant γ. Note that if $\epsilon < \gamma/T$ this will in general be a coarser mesh than $\mathcal{M}$.

2. Construct the quantum state corresponding to the output of the solver, as a superposition of basis functions from $\mathcal{M}'$.

3. Map this quantum state to the equivalent quantum state on the finer mesh $\mathcal{M}$. This is essentially equivalent to the classical task of expressing each element of $\mathcal{M}'$ in terms of elements of $\mathcal{M}$.

Here we are assuming that the meshes $\mathcal{M}$ and $\mathcal{M}'$ are sufficiently regular that the last step makes sense (in particular, that $\mathcal{M}$ is a submesh of $\mathcal{M}'$).

If $\epsilon \geq \gamma/T$, the state produced by the original subroutine is left essentially unchanged. If $\epsilon < \gamma/T$, the original state produced was within distance $O(1/T)$ of the actual solution to the corresponding FEM instance, as is the state produced by the new subroutine. By the triangle inequality, the new state must be within distance $O(1/T)$ of the old state. If each such state produced by one of the new subroutines is within Euclidean distance $O(1/T)$ of the corresponding original state produced by one of the QLE subroutines, then using a similar argument to § 5.8.1, the whole algorithm does not notice the difference between the original and modified sequence of subroutines except with low probability.

We now examine the complexity of the steps in the modified subroutines. Each use of step 1 solves the FEM with precision $O(1/T)$, which requires time $\text{poly}(T)$ and a mesh of size $\text{poly}(T)$. In step 2 we need to construct a known $\text{poly}(T)$-dimensional quantum state. This can be done in time $\text{poly}(T)$ for any such state (see e.g. [186, Claim 2.1.1]). If $\mathcal{M}$ and $\mathcal{M}'$ are suitably regular, the mapping required for step 3 can

be implemented efficiently, i.e. in time polynomial in n, the number of qubits used by the original algorithm.

As the original quantum algorithm solved T instances of the FEM, and acts nontrivially on all n qubits, its runtime must be lower-bounded by $\max\{T, n\}$. Therefore, the runtime of the new algorithm is at most polynomial in the runtime of the old algorithm. As the new algorithm no longer contains any quantum subroutines which solve the FEM, we see that any quantum speedup achieved by it does not come from quantum acceleration of the FEM.

5.8.3 Solving oracular FEM instances

We finally observe that there cannot be an efficient quantum (or classical) algorithm for solving an instance of the FEM if the input function $f(x)$ is initially unknown and provided via an oracle ("black box"), and does not satisfy some smoothness properties. Indeed, this even holds for near-trivial FEM instances.

Imagine we are given an FEM instance of the form $u(x) = f(x)$, for $f \in L^2[0,1]$, and are asked to approximate the quantity $\int_0^{\frac{1}{2}} u(x)^2 dx$ to within accuracy ϵ – this is a very simple property of a trivial PDE. Further assume that we are given access to f via an oracle which maps $x \mapsto f(x)$ for $x \in [0,1]$, and that there are N possibilities for what the function f can be. We will show that this problem is hard by encoding unstructured search on N elements as an instance of the FEM.

Let B be the "bump" function defined by $B(x) = \exp(-1/(1-x^2))$ for $-1 < x < 1$, and $B(x) = 0$ elsewhere. Fix N and let f_0 be the shifted and rescaled bump function $f_0(x) = \sqrt{N}B(2Nx-1)$. f_0 is supported only on $[0, 1/N]$, has continuous derivatives of all orders, and $\|f_0\| = \Theta(1)$.

Assume we have access to an oracle function $O : \{0,\ldots,N-1\} \to \{0,1\}$ such that there is a unique $y_0 \in \{0,\ldots,N-1\}$ with $O(y_0) = 1$. It is known that determining whether $y_0 < N/2$ or $y_0 \geq N/2$ requires $\Omega(\sqrt{N})$ quantum queries to O [97]. We define f in terms of O as follows. Given $x \in [0,1]$, set $y = \lfloor Nx \rfloor$ and evaluate $O(y)$. If the answer is 1, return $f_0(x - y/N)$. Otherwise, return 0.

f (equivalently, u) is a bump function on the range $[y_0/N, (y_0+1)/N]$, and is zero elsewhere. So, if $y_0 < N/2$, $\int_0^{\frac{1}{2}} u(x)^2 dx \geq C$ for some constant $C > 0$, while if $y_0 \geq N/2$, $\int_0^{\frac{1}{2}} u(x)^2 dx = 0$. Hence approximating this integral up to additive accuracy ϵ, for sufficiently small constant $\epsilon > 0$, allows us to determine whether or not $y_0 < N/2$. As this task requires $\Omega(\sqrt{N})$ quantum queries, solving this instance of the FEM must require $\Omega(\sqrt{N})$ queries to f. A similar classical lower bound of $\Omega(N)$ queries also holds. Note that this does not contradict the bound in Eq. 5.20 as the norms of derivatives of u are large.

5.9 Outlook

We have shown that when one compares quantum and classical algorithms for the FEM fairly by considering every aspect of the problem (including the complexity of producing an accurate approximation of the desired classical output), an apparent exponential quantum advantage can sometimes disappear. However, there are still two types of problem where quantum algorithms for the FEM could achieve a significant advantage over classical algorithms: those where the solution has large higher-order derivatives, and those where the spatial dimension is large.

For ease of comparison with the quantum algorithm, we have only considered a very simple classical FEM algorithm here; there is a large body of work concerned with improving the complexity of such algorithms. For example, the finite element mesh can be developed adaptively and refined near parts of the domain which are more complex or of particular interest. This can substantially improve the convergence speed. It is our suspicion that more advanced classical FEM algorithms might eliminate the quantum algorithm's advantage with respect to BVPs whose solutions have large higher-order derivatives. For example, adaptive schemes such as "hp-FEM" have, in principle, a discretisation error that scales far better than the scaling shown here; it can be shown [99] that a perfect adaptive scheme has scaling

$$\|u - \tilde{u}\| = O(e^{-1/h}), \tag{5.81}$$

provided that the dimension of the domain is both small and fixed. While this is a large improvement over the "vanilla" classical complexity presented above, it is not always apparent how to generate adaptive schemes that are effective enough to saturate this scaling, in practice. Also, it does not seem impossible that the quantum algorithm could be substantially improved using similar adaptive schemes.

Additionally, the possibility of substantial improvement in the classical algorithm is less clear with respect to problems in high spatial dimension d. Indeed, any reasonable discretisation procedure seems likely to lead to systems of linear equations which are of size exponential in d (the so-called "curse of dimensionality"). This is precisely the regime in which the quantum algorithm might be expected to have a significant advantage. One setting in which such high-dimensional BVPs occur is mathematical finance; for example, the problem of pricing multi-asset basket options using the Black-Scholes equation [122]. Alternatively, producing a solution to any problem in many-body dynamics requires solving a PDE where the dimension grows with the number of bodies. However, Monte Carlo methods can sometimes be used to alleviate the curse of dimensionality in practice [29, 136]. It is thus a compelling open question whether quantum algorithms can yield an exponential speedup for problems of practical interest in this setting.

CHAPTER 6

SIMULTANEOUS TESTING OF QUANTUM MECHANICS AND GENERAL RELATIVITY WITH A QUANTUM OPTICAL SATELLITE

In this chapter we deal with a tangential notion of validation of quantum systems: we propose an experiment designed to validate that a specific general relativistic system also behaves quantum mechanically, via the observation of a general-relativistic effect on single photon interference. The experiment consists of a folded Mach-Zehnder interferometer, with the arms distributed between a single Earth orbiter and a ground station. By compensating for other degrees of freedom and the motion of the orbiter, this setup aims to detect the influence of general relativistic time dilation on a spatially superposed single photon. The chapter details a payload to measure the required effect, along with an extensive feasibility analysis given current technological capabilities.

6.1 Introduction

Since the beginning of the 20^{th} century, quantum mechanics and general relativity have been the theoretical foundations of modern physics. However, while their predictive power has enabled us to understand the universe at both very small and very large scales, their validity has only been explored separately. Regarding general relativity, historic tests include measuring the perihelion of Mercury [62] and the Pound-Rebka experiment to quantify gravitational redshift [184]. Even with the precision attainable with contemporary experiments, general relativity remains robust; for example, estimates of the Shapiro delay (the slowing of light as it passes by a massive body) have been accurately measured via radio links with the Cassini spacecraft [22]. Additionally, tests of the (weak) equivalence principle now concur with general relativity with accuracy 10^{-13} [231], and will be bounded by 10^{-15} with

the space-based MICROSCOPE experiment [74]. This resilience to experimental analysis is also shared by quantum mechanics. The most stringent tests of quantum theory are precision tests of quantum electrodynamics (QED); these have now been verified to an accuracy of 10^{-12} using a one-electron quantum cyclotron [172].

While quantum field theory, which is currently the best description we have of the quantum world, is fundamentally incompatible with general relativity, descriptions of quantum field theories on static but curved background spacetimes are internally consistent and produce novel predictions [23] (such as Hawking radiation [107]). However, any full description of gravity itself as a quantum field theory is bound to fail due to a lack of renormalisability. Additionally, quantum field theory and general relativity have contrasting descriptions of physical parameters, the most notable of which is the nature of *time*; this "problem of time" is an unresolved hurdle in producing a canonically quantised theory of gravity [119].

Without exception, all of the experiments mentioned above are tests of either general relativity or quantum mechanics, but not the interplay of both. As such, their outcomes are consistent with classical dynamics evolving on a curved background (in the former case), or of a quantum field theory evolving in flat space (in the latter case). A controlled experiment exploring quantum dynamics on a curved background is, to the author's knowledge, yet to be performed. To this end, a theoretical proposal by Zych et al. [237] outlined the premise of a controllable experiment on a system that is both highly quantum, and occurring on a non-trivial background spacetime. Such an experiment would provide vital evidence to aid the successful unification of quantum mechanics and general relativity.

The crux of the proposal in [237] is to perform interferometry of single quanta, but to orient the interferometer in such a way that quantum states in each arm traverse a different gravitational equipotential. Historically, such an idea is not new; the analysis of gravitational effects on matter interferometry was first explored by Collela, Overhauser and Werner ("COW") in 1975 [66]. While the COW experiment did observe a gravitational influence on the output of the interferometer, their data is adequately described by analysing a Mach-Zehnder interferometer placed in a *Newtonian* gravitational field, rather than a truly relativistic spacetime. The key physical extension of the COW experiment present in [237] is to construct an interferometer large enough that general relativistic effects become apparent in the output of the interferometer, and to let the position of the photon in the interferometer serve as the local clock. Similar proposals exploring gravitational effects on quantum optical systems consider interferometry with a precessing polarisation serving as the clock [36], or explore the degradation of entanglement due to gravity [37].

The objective of this chapter is to present a full design of a space-based quantum optics experiment capable of carrying out the experiment proposed by Zych et al. [237], along with a discussion of the practical feasibility of such an experiment. The design consists of a "folded" Mach-Zehnder interferometer distributed between an orbiting satellite (containing a single photon source, a spool of single-mode optical fibre, and a transmitting telescope) and a ground-station (equipped with a receiving telescope, an identical spool of fibre and a bank of single photon detectors). A stripped-down schematic of the proposed design is shown in Figure 6.1. Moreover, we contend that overcoming the engineering requirements for this proposal would be of significant benefit to those interested in developing quantum communications networks on a global scale.

The chapter breakdown is as follows: in § 6.2, we give a scientific background, and extend the analysis present in [237] to derive the predicted interferometric signal given that the satellite is in motion. In § 6.3, we give an extensive feasibility analysis of the experiment. This includes a proposed setup and operational parameters for key optical components in § 6.3.1, followed by analysis of interferometric stability, systematic error, loss and noise in § 6.3.2, § 6.3.3, § 6.3.4 and § 6.3.5, respectively. All of these elements are combined with a statistical analysis in § 6.4, in order to derive the operational duration necessary to test the hypotheses in [237] to a prespecified confidence. Finally, we outline orbit and mission requirements in § 6.5 and highlight operational risks to be mitigated in § 6.6.

6.2 Scientific background

While the experiment proposed by Zych et al. in [237] calculates the general relativistic effect on the fringe visibility of the interferometer output signal, we must also introduce special relativistic corrections due to the orbital velocity of the satellite (§ 6.2.1). This full expression for the time dilation is then used to derive an expected output for the Mach-Zehnder interferometer in a regime where relativistic corrections are important (§ 6.2.2).

6.2.1 Relativistic time dilation

General relativity predicts that a clock in a lower gravitational potential ticks slower than a clock in a higher one. This means that two clocks initially synchronised and then moved to points of differing gravitational potential accumulate an increasing time delay. This gravitational time dilation is observed daily in global navigation satellite systems, where one also has to account for special relativistic time dilation due to the motion of the positioning satellites. The total time delay can be derived

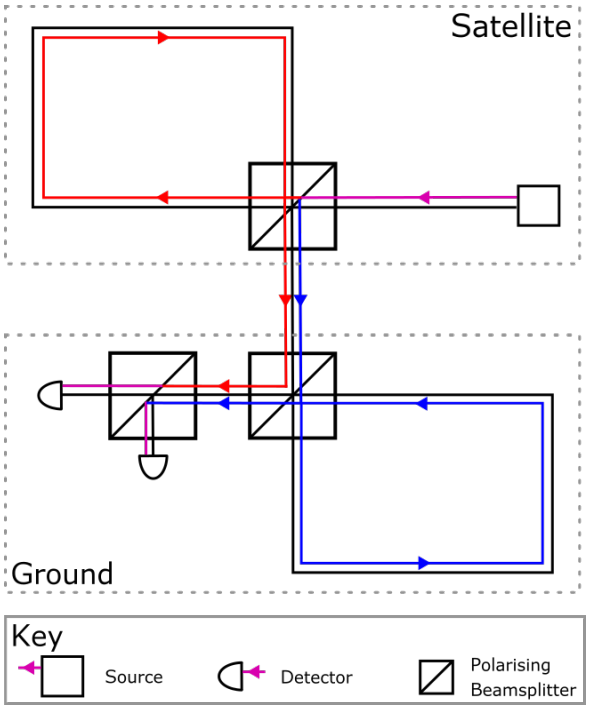

Figure 6.1: A simplified schematic of the proposed experiment. Colour here does not represent frequency; red and blue denote orthogonal polarisation states. The red and blue paths are only distinguished by their travel times around the loops on the satellite and on the ground, respectively.

directly in a general relativistic framework starting from the spacetime metric, using the weak field approximation for the Earth's gravitational potential and assuming the speed of a satellite to be much smaller than the speed of light. It is important to note that the time dilation is independent of the physical realisation of the clock. For example, we could consider an optical fibre loop and use the position of a photon travelling in the fibre as a clock. If two of these fibre-clocks were placed at different heights in the Earth's gravitational field, the one at the lower position would tick slower, meaning that the photon would take more time to travel the entire length of the fibre. Such a time delay is known as the *Shapiro delay*.

Considering the simplified experimental setup described in Figure 6.1, we deduce that the time spent by a photon in the satellite loop differs from the time spent in the ground loop, as measured by an observer on the ground. The accumulated time delay, i.e. the difference in photon arrival times, can be obtained by generalising the calculation described in [237] (see Appendix D.1). For a satellite on an elliptical orbit around the Earth, the time delay is given by

$$\Delta\tau = \frac{nl}{c^3}\left[-W_0 + GM\left(\frac{2}{R_\oplus + h} - \frac{1}{2a}\right)\right] - \frac{ndl}{c}, \tag{6.1}$$

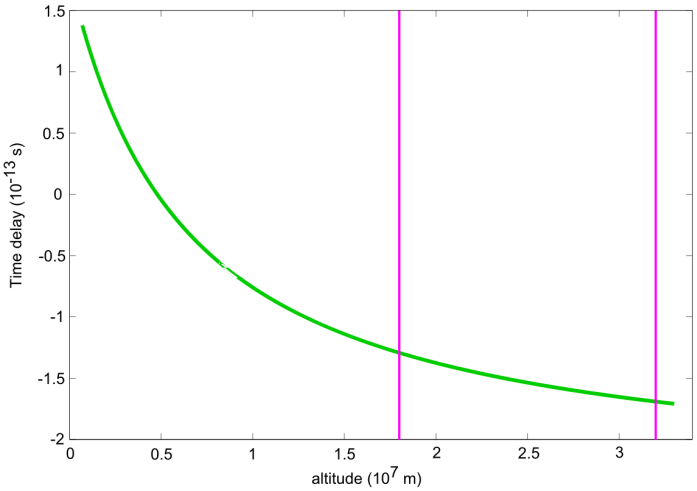

Figure 6.2: The total time delay between the fibre clocks, as a function of height, for a fixed fibre length and fibre refractive index. Magenta lines mark lowest and highest altitudes at which measurements could be performed (see feasibility analysis). In this region the time delay is dominated by the Shapiro delay. The choice of fibre length and refractive index is discussed in § 6.3. The altitudes and orbital parameters are discussed in § 6.5.

where l is the proper length of the satellite fibre, $l + dl$ is the proper length of the ground fibre, n is the fibre's refractive index, G is the gravitational constant, M is the mass of the Earth (5.97219×10^{24} kg), $R_\oplus$ is the radius of the Earth (6378100 m), the geoidal potential $W_0 = L_G c^2$ where $L_G = 6.969290134 \times 10^{-10}$ by definition, c is the speed of light in vacuum, h is the altitude of the satellite with respect to the Earth's surface and a is the semi-major axis of the orbit. In order to compute the gravitational potential experienced by the satellite, we assume the Earth is a perfect sphere. We can safely neglect the corrections due to the irregular shape of the Earth, both on the gravitational potential and on the satellite's orbit, since they are orders of magnitude smaller than the first order approximation [130]. A plot of this expression, in the case $dl = 0$, is shown in Figure 6.2. Note that the general relativistic and special relativistic time dilations act in opposition, such that at low satellite altitudes the time delay is positive and dominated by the special relativistic dilation, whereas at high altitudes it is negative and dominated by the general relativistic dilation. Therefore care must be taken to launch the satellite into a sufficiently high orbit for the general relativistic time dilation to be the dominant effect. Figure 6.2 demonstrates that the Shapiro delay for a satellite at the orbital heights considered in this proposal is of the order 100 fs.

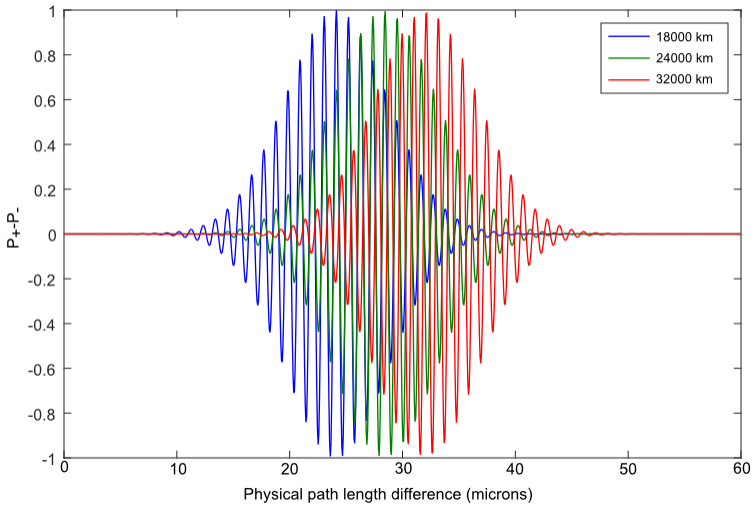

Figure 6.3: The expected interferometer signal vs. physical path length difference of the two interferometer arms, with the parameters as described in the text, at three different heights of the satellite. The signal was calculated by numerically evaluating Eq. (6.2) with a frequency-dependent refractive index as described in [150].

6.2.2 Interferometry in the presence of gravity

Given this modification to the expected time dilation, we provide a brief exposition on the theory presented in [237], in order to motivate the design of our experiment. The reader is referred to [237, §2, §3] for a more thorough treatment.

The output of a Mach-Zehnder interferometer is sensitive only to a difference in the relative optical path length of each arm [144]. If the interferometer were to have two ideally-controlled arms of equal proper length (i.e. $dl = 0$), the relative optical path would be dependent only on the Shapiro delay $\Delta\tau$. However, by introducing a controlled difference in the relative length of the two interferometer arms by actively controlling the length of the fibre spools, one could cancel out $\Delta\tau$ and recover maximal contrast at the output of the interferometer for any height of the satellite. Measuring the path-length difference that gives maximum contrast at different gravitational potentials would constitute a measurement of the Shapiro delay between the parts of the photon state traversing the upper and lower arms of the interferometer. It can be shown (see Appendix D.2) that the probability of the photon being detected at the $\pm$ outputs of the Mach-Zehnder interferometer (with a single photon input) is given by:

$$P_{\pm} = \frac{1}{2}\left(1 \pm \int d\nu |f(\nu)|^2 \cos(\nu\Delta\tau)\right), \tag{6.2}$$

where ν is angular frequency in rads^{-1} defined in the reference frame of the observer at distance r from the Earth, $\Delta\tau$ is the time delay as given in Eq. 6.1, $P_{\pm}$ is the

probability of the photon being detected at the ± output, and $f(\nu)$ is the spectrum of the light source, normalised such that $\int_{-\infty}^{\infty} |f(\nu)|^2 d\nu = 1$.

In Figure 6.3 we numerically evaluate Eq. 6.2 with a range of different fibre lengths for three different satellite heights. We assume a Gaussian spectral density for the light with $f(\nu) = \left(\frac{\sigma}{\pi}\right)^{1/4} \exp(-\frac{\sigma}{2}(\nu - \nu_0)^2)$, with parameters taken from the setup in § 6.3: a central frequency $\nu_0 = 2\pi \times c/1550\,\text{nm}$ and width $\sigma = (100\ \text{fs}/2\pi)^2$. We take the satellite fibre to be 60 km long, while the length of the ground station fibre differs by the amount given on the x-axis of the figure. The frequency-dependent speed of light in the fibres is calculated from the refractive index for fused silica given in [150]. In this proposal, a satellite is sent on an elliptic orbit and we record the physical path-length difference required to recover full visibility, for varying satellite heights. From this, the corresponding Shapiro delay as a function of satellite height can be inferred.

6.3 Implementation

In order to present a feasible implementation, it must be demonstrated that enough data can be collected within the lifetime of the satellite to confirm, or refute, the hypothesised signal to a prespecified confidence. In this particular proposal, there are four key issues that must be addressed before feasibility can be demonstrated: stability of the interferometer, loss from the source to the detectors, noise from extraneous and atmospheric photons, and mitigation of any other degrees of freedom besides the Shapiro delay that might introduce distinguishability (in particular, photon dispersion). We first outline a feasible setup in § 6.3.1, before presenting an analysis of these key issues.

6.3.1 Optical setup and components

A detailed illustration of the interferometer setup, with all the key optical components, can be seen in Figure 6.4. The choice of operational parameters is ultimately a compromise between what is feasible to implement and what is necessary to recover the desired data. As such, there is no single optimal set of components; relaxing the constraints on their performance may be acceptable if they become more durable and the satellite can operate for longer and take more data, for example (we aim to mitigate large pulse dispersion, noise and random fluctuations in signal by integrating counts over long time periods; hence, durability of components is critical). Therefore, the following choices of components and parameters are not intended to be a fixed specification for the payload, but to

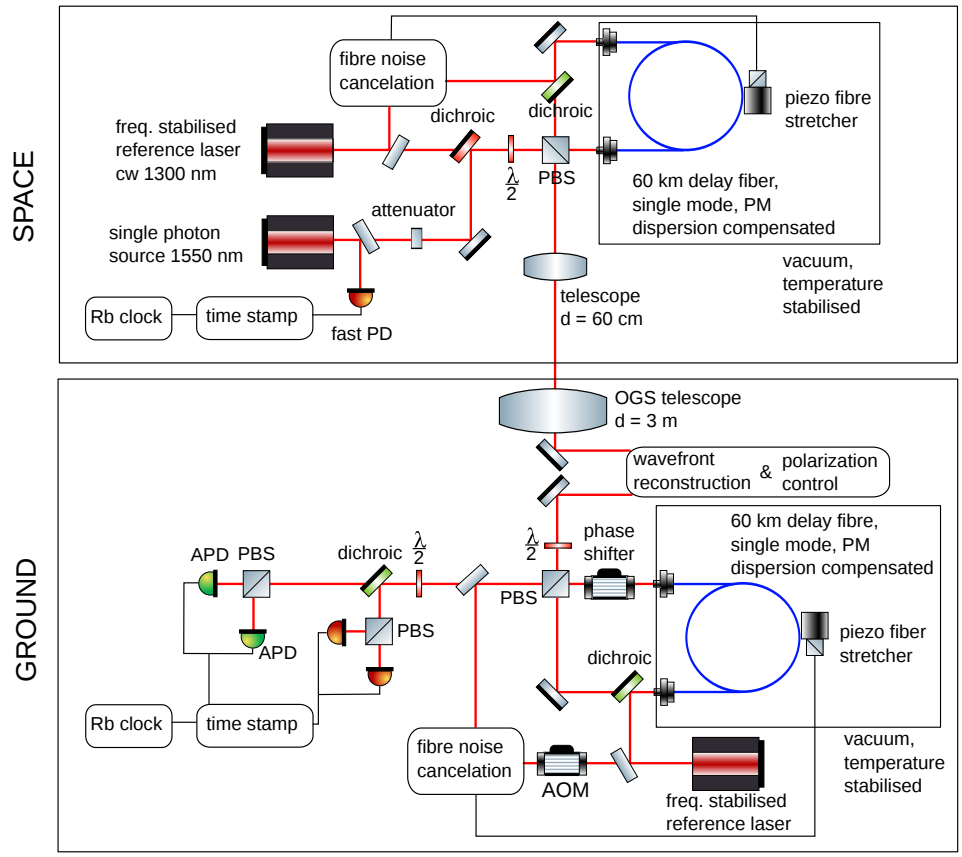

Figure 6.4: Full schematic of the payload optical system. The setup is designed to minimise the effects of other degrees of freedom besides the Shapiro delay that may introduce distinguishability at the output of the interferometer.

demonstrate one setup that could feasibly produce meaningful results. We highlight the following optical elements, whose operational parameters are crucial to the success of the experiment.

Single photon source: An off-the-shelf, mode-locked, pulsed laser with operational wavelength of 1550 nm and 1 GHz repetition rate is attenuated to a mean photon number of 0.1 photons per pulse. While an attenuated laser is a straightforward option, true, heralded single photon sources have been space-qualified and operated on the QUESS satellite [88]. However, we are not aware of a true, heralded source of single photons that meets the requirements of repetition rate and pulse width presented here. The reason for this particular level of attenuation is the trade-off between lowering multiphoton emission and maintaining a high number of total counts - this attenuation reduces the probability of multiphoton emission to 5% and results in a single photon rate of 100 MHz, which matches the resolution of the single photon detectors. The chosen wavelength

guarantees high atmospheric transmission and the utilisation of well-developed telecommunication technology. The functional dependence of the fringe contrast on the Shapiro delay $\Delta\tau$ and the pulse width $\sqrt{\sigma}$ (see [237], Eq. 13) demands the utilisation of ultra-short (<1 ps) single photon pulses to measure a noticeable effect (we note single photon sources with a width as short as 65 fs have been demonstrated in [166]). We take $\sqrt{\sigma} = 150\,\text{fs}$.

Classical reference laser: A 1300 nm multi-purpose laser, operating in continuous wave mode, is led alongside the single photons through the interferometer. Its purpose is to provide reference data to estimate phase fluctuations and systematic errors. For precise corrections, an operational wavelength as close as possible to the single photon source is required, but overlap with the bandwidth of the single photon pulses has to be avoided. The chosen wavelength is a reasonable compromise. Frequency stability of both lasers is paramount; see § 6.3.2. Additionally, a fraction of the incident laser power is separated at the ground station and used for wavefront reconstruction and compensation of polarisation changes due to satellite movement.

Fibres: We expect the delay fibres must be stabilised to relative length changes of 10^{-10} for observing the predicted interference effects, which includes a thermal stabilisation of $\pm 10^{-5}\,\text{K}$. Moreover, the fibre length must change dynamically to recover full contrast of the interference fringes. Active length corrections are carried out by a piezo fibre-stretcher. Fibres of length $l = 60\,\text{km}$ and refractive index $n = 1.5$ (glass) are assumed.

Transmission telescope: The on board emitting telescope is used to focus the beam from the two sources. The aperture needs to be large enough to emit a strong signal but is limited by size and weight; a reasonable choice is a 60 cm aperture, 1 m long, 6 μrad field of view, Cassegrain reflector telescope. Material choices, such as Beryllium mirrors, would also minimise weight.

Single photon detectors: Since the goal of the detectors is to receive single photons that have travelled astronomical distances, it is extremely beneficial to minimise dark counts and maximise efficiency. We consider single-nanowire single photon detectors (SNSPDs), to benefit from the superior dark count over conventional avalanche photodiodes. While these detectors aren't as commonplace, the technology has already been established in a similar setting in the LLCD mission (NASA) [24].

Timing: Emission of the single photon pulses are tagged with timestamps by a Rb-clock (chosen for its small size, weight and commercial availability). Synchronisation with an identical clock on the ground will be used to exclude background noise at the data post-processing stage and to track the path of the

satellite. Precision timing is also necessary to modulate the action of time-gate filters in front of the detectors.

6.3.2 Random fluctuation and stability

As the effect we wish to measure is a minute change in optical path length in the interferometer due to the Shapiro delay, exposing the interferometer to other sources of instability can swamp the desired signal. We find that optical path and phase changes in the atmosphere due to temperature fluctuations are negligible compared to fluctuations in the length of fibre; given the thermal properties of fused silica we calculate necessary temperature stabilisation of the fibres of less than 10^{-5} K to ensure a relative path length stability of 10^{-10}. Passive insulation and active heat distribution can mitigate a large fraction of the thermal instability, but both the satellite and the ground station include a feedback loop of a frequency stabilised reference laser in combination with a piezo fibre stretcher, to allow for fibre noise reduction. Also, some thermal fluctuation can be erased in post-processing by referring to data from the reference laser.

In addition, continued operation of the experiment requires protection of the reference laser from frequency fluctuations and drift. Given the magnitude of the path length change, we estimate a necessary relative frequency stability of the reference laser less than 10^{-11}. Long term accuracy of the reference laser within required precision is most feasible via frequency comb stabilisation [80]. This method will increase further complications and costs, and could be circumvented by improvements in the area of stabilisation by using atomic absorption lines, which today are close to reaching comparable relative accuracies [121].

6.3.3 Systematic transmission errors and dispersion

We highlight three systematic errors present in the experiment, mitigation of which are critical: dispersion, both in the fibre and the atmosphere; Doppler shift due to the velocity of the satellite; and changes in optical path length due to the ellipticity of the orbit. Dispersion is prevalent both in the optical fibre and in the atmosphere. We estimate a requirement for fibre dispersion of < 5 fs/km/nm, ensuring a broadening of the pulse width in the fibres of $< 0.5\%$ per km; this is a stringent enough requirement to fix dispersion as the primary hurdle facing this experiment. The current state-of-the-art for dispersion-limited fibre is a factor of ten worse than this: 50 fs/km/nm [204]. Current technologies are capable of dispersion compensation in optical fibres of 0.5 ps/km/nm, which has to be improved by about a factor of ten to make the scientific requirement feasible. However, the utilisation of telecommunication fibre wavelengths give the greatest chance of breakthrough

research in that scientific area. For example dispersion-free transmission of 610 fs laser pulses over a 160 km fibre has been demonstrated, but dispersion-free transmission techniques typically suffer from large losses [216, 180]. On the other hand active compensation of *atmospheric* chromatic dispersion of a 250 fs pulse over a 200 km propagation distance, to a uncertainty of ±10 fs, has been demonstrated [137]. This indicates that control of atmospheric dispersion of comparably short pulses is well within reach, especially considering the pulses in this experiment are travelling vertically, on a much shorter trajectory through the atmosphere. Doppler shift, conversely, presents much less of a problem. This is straightforward to calculate based on the velocity of the satellite relative to the Earth. For the worst possible case (at apogee), the Doppler redshift v is calculated using the formula:

$$\Delta\nu_{Doppler} = f_p \times \frac{c}{c - v_0} \tag{6.3}$$

where c is the speed of light, f_p is the frequency of the single photon source and v_0 is the maximum velocity of the satellite at the lowest sampling point. This gives a shift in photon frequency of 2.48 GHz. Considering the bandwidth size and that the photon frequency exceeds 193 THz, this is considered negligible. Finally, the radial motion of the satellite causes constant variation in optical path length. To keep the path lengths of the interferometer equal this variation must be continuously compensated for. The time spent in the fibre loop by the part of the photon state in the upper arm, before it is emitted from the satellite, is given by:

$$\Delta t = \frac{\ell}{cn} \approx 10^{-4}\,\mathrm{s} \tag{6.4}$$

where $\ell = 60$ km is the fibre length and $n = 1.5$ is the refractive index. Multiplying this quantity by the radial velocity gives the change in path length between parts of the superposition due to the motion of the satellite; this change in path is plotted in Figure 6.5. However, this effect can be precisely calculated in advance and so either active compensation with optical components or passive compensation at the post-processing stage can be built into the experiment beforehand.

6.3.4 Loss

Given the repetition rate of the laser and strength of attenuation, we can calculate the expected rate of signal photons registering at the detectors. The analogous discussion on noise photons reaching the detectors is deferred to § 6.3.5. There are three major sources of signal attenuation: divergence of the beam from the transmitting telescope to the ground; loss due to atmospheric irradiance and interference; and loss within the fibres.

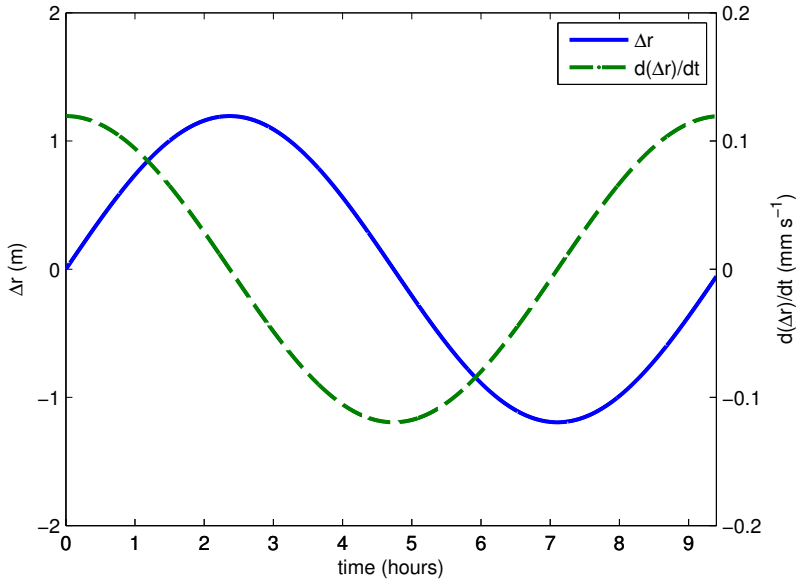

Figure 6.5: Path length difference and velocity difference vs. time, due to motion of the satellite. The solid blue line shows the free-space path length difference, and the dashed green line shows the rate of change of this difference. This is compensated for in post-processing, rather than in-flight.

First, we examine the effects of atmospheric transmission. A wavelength of 1550nm is deemed suitable as it is standardised for telecoms use, maximises transmittivity in the fibre and is not readily absorbed by the atmosphere. The high frequency of the emitted photon allows us to ignore ionospheric effects on polarisation and is easily distinguishable from auroral activity over the ground stations [46]. Assuming homogeneity, using a variant of the Beer-Lambert Law we can estimate the probability of a single photon passing through the atmosphere as:

$$\Pr(\text{transmission}) = \exp\left(\frac{-\tau_0}{\eta_0}\right) \tag{6.5}$$

Where τ_0 is the optical depth of the atmosphere, and η_0 is the angle of incidence of the beam. The optical depth varies over time and depends on myriad dynamic factors such as aerosol content, atmospheric mixing and Raman and Rayleigh scattering. Modelling these effects is essential, so one might consider optical depth readings from MODIS, MISR or future Sentinel satellites. A numerical weather prediction model might then be produced detailing daily optical depth over ground stations. An alternative might be to model turbulence transfer functions as in [195]. As shown in Figure 6.6, we can allow for an optical depth of 0.5 before the satellite signal is severely attenuated. Combining these factors produces an estimate for the atmospheric transmission loss to be $\simeq -2.3\,\text{dB}$.

As for beam divergence, a downlink direction is chosen as accentuated divergence

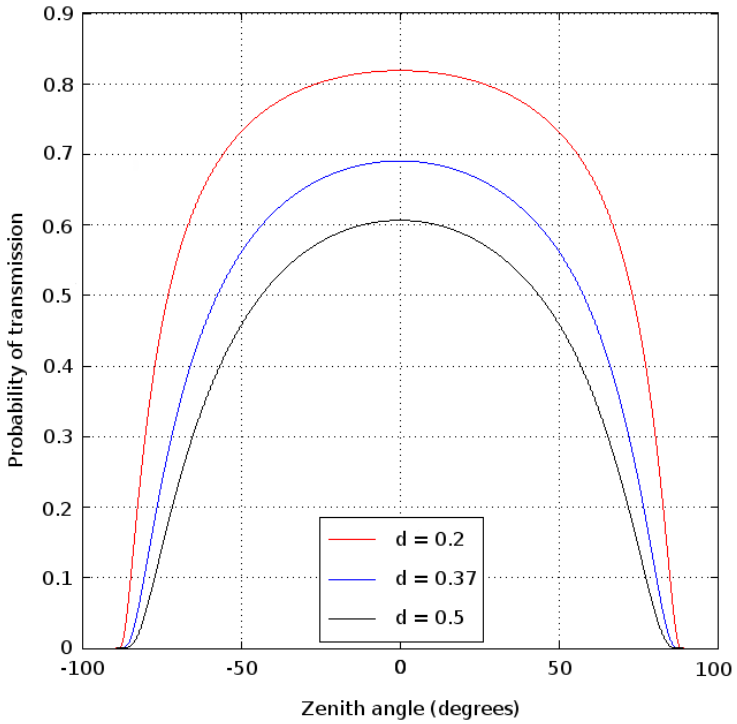

Figure 6.6: Probability of a photon to be transmitted through the atmosphere with respect to zenith angle, θ, and various optical depths, d. An optical depth, d, of 0.5 corresponds to high levels of pollution. Further calculations assume d=0.37, giving an atmospheric transmission probability of 0.69 at zenith.

would then occur only in the last 12 km of travel. This downlink beam divergence can be computed for a photon of wavelength λ by:

$$\theta_{div} = \frac{\lambda}{\pi w_0} = 1.6\,\mu\text{rad} \tag{6.6}$$

where $w_0 = 30\,\text{cm}$ corresponds to the radius of the transmitting telescope on the satellite. Consequently, the beam diameter on the ground at maximum and minimum altitude is given by $D_{ground} \simeq 2h\theta$, which implies that $58\,\text{m} \leq D_{ground} \leq 105\,\text{m}$, taking bounding values for the altitude as $18000\,\text{m} \leq h \leq 32000\,\text{m}$.

Atmospheric turbulence must also be considered. As the atmospheric parameters used for post-processing are extracted from data from the reference laser, this turbulence must not radically change in the time between measurement of the reference laser and of the single photon source. Assuming that the gap between reference and single photon measurement is half the repetition rate of the laser, this delay comes to 5 ns. Assuming a wind speed of $10\,\text{ms}^{-1}$, an air parcel would move by about 50 nm in some direction between the single photon

measurement and reference laser measurement. This is considered negligible compared to other atmospheric effects.

Atmospheric transmission loss, turbulence and beam divergence can be compiled into a "link budget" that calculates the full transmission loss from satellite to ground station. We can modify and simplify the Friis transmission equation to give the following link budget equation [203]:

$$\text{Loss}(dB) = 10\log\left[\left(\frac{\pi D_T D_R}{4\lambda h}\right)^2 L_p L_t\right], \tag{6.7}$$

where here D_T is the transmitter diameter, D_R is the receiver diameter, λ is the wavelength of the single photon source, h is the satellite altitude, L_p is the pointing loss (taken to be 0.63 [215]), and L_t the atmospheric transmission loss as calculated above. Assuming an apogee of 32000 km and a lowest altitude bin of 18000 km, a ground receiver diameter of 3 m, a satellite transmitter diameter of 0.6 m, and superconducting detector efficiencies of about 90% [152]; we calculate the baseline signal gain from the link budget to be -30 dB at perigee, and -35 dB at apogee. Further factoring in losses from the optical fibre of -15.5 dB and fibre blackening from radiation of -1.7 dB, the link budget achieves a total attenuation of -52.2 dB in the worst case, towards the end of mission lifetime. Of course, to further reduce signal attenuation, one could increase the aperture diameters of the receiver and transmitter. However, this causes a loss in manoeuvrability and a significant cost increase for rapidly decreasing returns.

6.3.5 Noise

A feasibility case must also ensure a signal-to-noise ratio (SNR) great enough to produce enough meaningful data for statistical analysis. It must be stressed that the signal received on Earth is fixed by the link budget above, however since we are trying to detect a single photon from a plethora of solar and planetary photon noise, optimising the SNR is crucial.

The effect of noise on space to ground quantum channels has previously been explored [143]. The noise power received by the ground telescope (P_b) can be expressed as

$$P_\text{b} = \frac{1}{4} H_\text{b} \Omega_\text{fov} \pi D_R^2 \Delta\nu \Delta t_d\,, \tag{6.8}$$

where H_b is the brightness of the sky in units of $\text{W}\,\text{m}^{-2}\,\text{sr}^{-1}\,\mu\text{m}^{-1}$, Ω_fov is the field of view, $\Delta\nu$ is the bandwidth and Δt_d is the detection time. Typical sky brightness for quantum cryptography applications is discussed in [143], using data from [138]. However, the data therein must be modified in light of the experiment proposed

here. Firstly, the data presented in [143, 138] is for a frequency band just below 1550 nm, which is subject to much less noise than at 1550 nm itself. Conversely, the primary source of noise photons at 1550 nm is hydroxyl airglow, the strength of which is strongly dependent on ambient temperature. The data from [138] assumes a receiving station in the tropics at Mauna Kea, whereas we propose an arctic station at Svalbard. Utilising instead polar sky brightness data from [181], we take a sky brightness of $2\times10^{-5}\,\mathrm{W\,m^{-2}\,sr^{-1}\,\mu m^{-1}}$ at our operational frequency. The received signal band is filtered to a bandwidth of 15 nm, corresponding to twice the bandwidth of the single photon pulses (twice the full width at half maximum of the Lorentzian pulse). The field of view of the receiving telescope is assumed to be $10\,\mu$rad, but the effective field of view is further reduced by a factor of 10 with a variable iris diaphragm [200]. Noise power is further reduced with a 50 ps time gate filter leading to an available detection time of 50 ms per second. Reflections from the satellite or its black body radiation do not significantly contribute to the background noise [143]. Besides received background photons, total noise power also depends on the dark count rate of the single photon detectors. Superconducting detectors have negligible intrinsic dark count rate but as a worst case estimate the detector system dark count rate is assumed to be 1 kHz (although a large fraction of these counts are neglected due to time gating) [152]. The rate of detected signal photons is calculated as 590 Hz (using the repetition rate of the laser, attenuation and loss). Combining this figure with the expected noise from ambient photons and system dark count gives an SNR ≈ 9.0 =9.6 dB.

6.4 Hypothesis testing

Once an experimental profile has been outlined, we can perform a statistical analysis to extract the confidence with which we can confirm, or refute, the hypothesis in Eq. 6.2 (and likewise, the theoretical analysis presented in [237]). We assume data taken in k altitude bins, uniformly spaced in satellite height h from $h_{min} = 18000$ km to $h_{max} = 32000$ km, with n counts per bin. Then, we take normally distributed experimental parameters with mean as given in the specification in § 6.3.1, and variance as fixed by the discussion on stability in § 6.3.2. We assume that a fraction of the photons are lost according to the link budget in § 6.3.4, and that there is an ambient background of noise photons as per § 6.3.5. We then perform Monte Carlo simulations for a range of number of altitude bins, k, and counts per altitude bin, n, to generate artificial data sets given this initial prescription. In order to extract a statistical significance from these simulations, we perform a Kolmogorov-Smirnov statistical significance test between the simulated

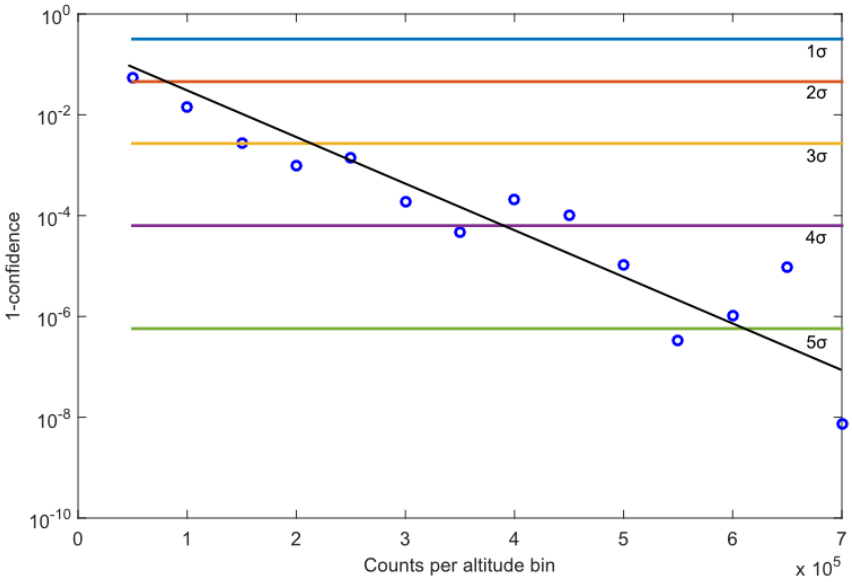

Figure 6.7: Monte Carlo estimation of the number of counts per altitude bin required in order to confirm (or refute) Eq. 6.2 to a specified confidence. Horizontal lines indicate increasing "sigmas" of confidence, descending from 1σ to 5σ.

data set and the mathematical formulation in Eq. 6.2 and [237]. The test is devised to extract the goodness of fit for a data set to an arbitrary functional hypothesis. Using this test, we can predict the number of single photons that must be sent in order to test the hypothesis to a specified confidence.

The consequential data from the simulation and statistical test is shown in Figure 6.7, for $k = 500$ altitude bins. Here, we can see that the number of received, true counts per altitude bin must exceed $\approx 6.5 \times 10^5$ counts in order to test the hypothesis to a $\gtrsim 5\sigma$ confidence. Given the figure for total loss of $-52.2\,\mathrm{dB}$, this amounts to a necessary emission total of $\approx 10^{11}$ single photons per altitude bin. Given the repetition rate of the laser and the degree of attenuation, we expect raw emission rates of 100 MHz and thus reach the required emission total in each altitude bin within ~ 1000 seconds $= 16.7$ minutes of continuous operation.

6.5 Mission design

In order to reach the desired confidence defined in § 6.4, data must be taken over a sufficiently large difference in gravitational potential. An orbit with a perigee of 700 km and an apogee of 32000 km gives access to a relativistic time delay of 150 fs, which is both large enough to be resolved by the detectors and still gives a reasonable count rate at the apogee. During the measurement procedure, every

other light source introduces noise. The ambient sunlight is indeed strong enough to wash out the signal from the satellite. Therefore, measurements must be performed when the ground station is not illuminated by the Sun. Additionally, in order to have the maximum number of measurements per orbit and to minimise noise photons from airglow, the ground station needs to be placed as close as possible to the North Pole. These considerations lead to the selection of the ground station located in Svalbard, Norway. By considering the maximum optical path that gives a valid measurement and the movement capabilities of the ground telescope, a connection cone of 45° around the zenith can be defined. This allows satellite access times for up to 7 hours per orbit. This ground station is ideal for the mission, though it introduces a constraint on orbital inclination. In fact, only a polar orbit can provide the maximum visibility time from the ground.

The range of orbital heights chosen to send single photons is from 18000 km to 32000 km (to minimise special relativistic effects, as discussed in § 6.2.1). Given the number of altitude bins and length of measurement windows, we estimate that a mission lifetime of 1.5 years will completely satisfy the previously stated requirements. The mission profile is thus composed of three different phases. The first, starting immediately after launch, is a 6-month commissioning period used to calibrate the orbit and the measurement system with laser ranging and radio communication. During the satellite motion, due to the perturbation from the oblateness of the Earth and the high eccentricity of the orbit, the orbital apse line rotates clockwise with a rate of 77.1°/year. With an initial argument of perigee of $\omega = 350°$ and a launch during Svalbard's spring season, after the commissioning phase $\omega \simeq 310°$, meaning that the apogee is contained in the ground station visibility cone. This situation is represented by the thicker line in Figure 6.8. The main operational phase thus begins in winter and lasts for approximately 6 months. It is then followed by a second mission phase in Svalbard's summer, during which the first results are examined and the orbit is further calibrated. Finally, the last 6-month phase provides additional measurements. At the end of the 1.5-year lifetime, a 55 ms^{-1} thruster burn lowers the perigee to approximately 200 km. This manoeuvre eventually leads to a further lowering of the apogee and then to a controlled de-orbit into the atmosphere.

The total satellite mass is approximated at 400 kg. The small spacecraft size allows using the VEGA launcher in order to bring the satellite into a parking orbit with a perigee of 700 km and an apogee of 20000 km. The final orbit is then reached using an on board bi-propellant propulsion system through a perigee burn with a ΔV of 323 ms^{-1} .

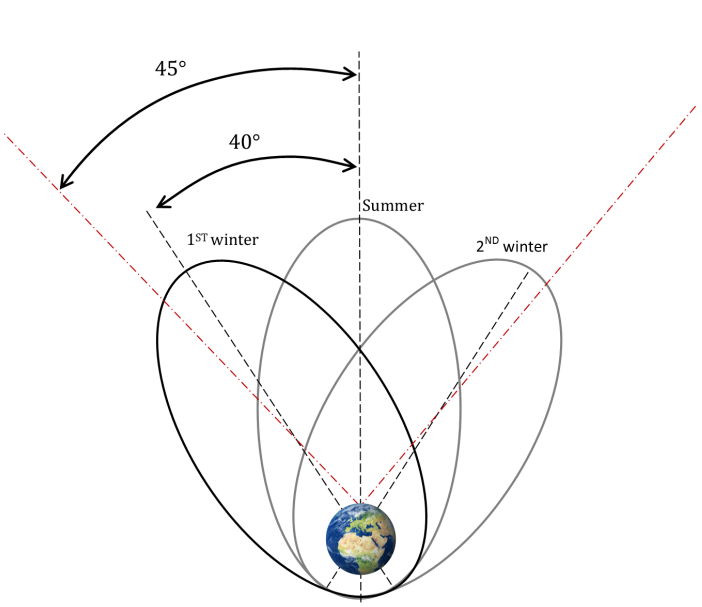

Figure 6.8: Schematic of mission phases and orbit drift. Note the division into first and second operational phases.

6.6 Risk analysis

6.6.1 Radiation effects

The spacecraft's immediate radiation environment, composed of fluxes from the solar wind and galactic cosmic rays, has a non-negligible effect on the performance of the components. During its orbit around Earth, the satellite passes twice through the Van Allen belts. Therefore, suitable shielding is necessary to ensure the correct operation of the payload. Particular precaution has to be taken to shield the spools of optical fibre, due to the effect of Radiation-Induced Attenuation (RIA). Single mode fibres with an RIA of 0.5 dB/km per 9000 rad have been demonstrated in the literature [221], a figure which can be used for a worst-case analysis. Assuming a 20 mm thick spherical layer of aluminium shielding, the fibre receives an estimated Total Ionizing Dose (TID) between 500 and 2000 rad/year, which yields an RIA between -1.7 and -6.7 dB/year. An aluminium shielding of 2 mm is provided for the lasers and the optical bench. This is also ample shielding to guarantee low radiation doses for the remaining optics. The blackening of fibre cables due to radiation is thus one of the main operational risks. Although this process is slowed by adequate shielding, occurrence at a higher than expected rate would pose a major obstacle to collecting adequate data. Active avoidance of the Van Allen belts by modifying the orbit could be pursued, if needed.

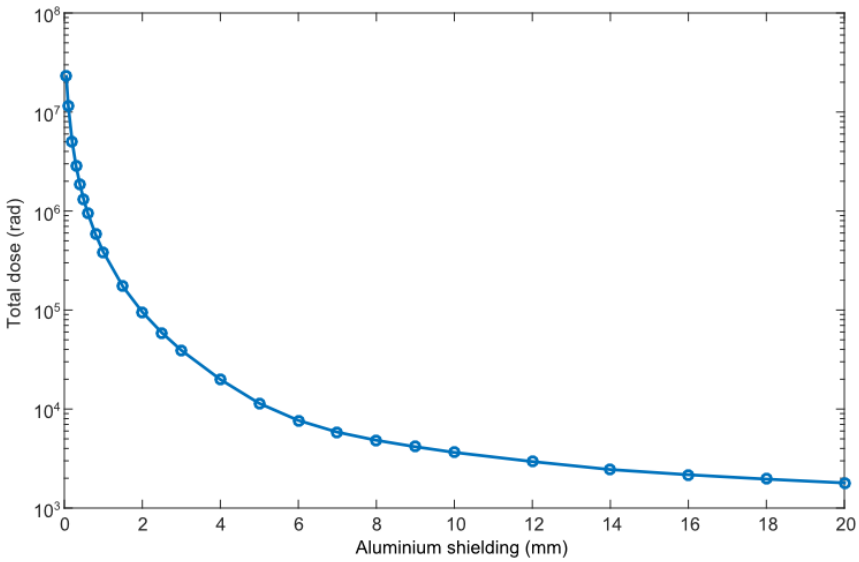

Figure 6.9: Total ionizing dose as a function of the aluminium shielding. Choice of shielding thickness is a compromise between continued performance of the optical fibres and payload weight.

6.6.2 Thermal control systems

During the mission, the satellite undergoes cyclic eclipses; the resulting temperature fluctuations have to be analysed in order to design the thermal control system. The most critical requirement is the temperature stability of the optical bench. Insulation ensures that the fibre temperature remains within 0.5 K of the equilibrium value. Specifically, the satellite is equipped with a Multi Layer Insulation (MLI) aluminium and Kapton coating and heat pipes to convey and redistribute heat. With the appropriate sizing for radiators and heat pipes it is possible to achieve the required thermal stability, although this represents another primary risk for the mission.

6.6.3 Attitude and orbit control systems

Precise determination of the satellite's orbit and fine pointing are also fundamental. Firstly, we assumed in § 6.3.4 that the satellite's pointing accuracy is 0.5 μrad, with 0.1 mrad/s maximum slew rate. This strict requirement ensures efficient transmission of the required number of photons to the ground, but is a relevant risk for the mission. Indeed, if the satellite failed on maintaining the required accuracy, a prohibitively large fraction of the signal photons would be lost. The satellite is thus equipped with a tracking telescope to have a first estimate of the attitude; then, two star trackers refine the determination. A system of reaction wheels and

vibration dampers allows for a fine control of the satellite orientation.

As mentioned above, orbital determination is of key importance for the mission. Indeed, in order to account for the transmission error due to the radial motion of the satellite, the radial velocity must be known to within $\sim 1\,\mathrm{mm/s}$ precision. This is feasible with current radio tracking systems working at either S- or X-band. In fact, the commonly used Ultra-Stable Oscillators often have a Van Allen stability better than 10^{-13} over an integration time from 10 s to 1000 s, which is completely adequate to fulfil the requirement.

6.7 Outlook

We have demonstrated herein a proposal capable of probing the interplay between quantum mechanics and general relativity using single photon interferometry - in particular, exploring both the dichotomous nature of time in the two theories, and the extent to which minimal coupling models are a good fit for a prospective theory of quantum gravity. Moreover, the apparatus is well within the reach of current quantum optics technology, and there is already historical precedent for space-qualification and launch of similar payloads. Conversely, we highlight optical pulse dispersion as a particularly acute problem for the successful operation of this experiment in the near future. However, if this hurdle can be overcome by near-term dispersion compensation devices, this experiment would itself provide a strong technological precedent for future projects involving commercial quantum communications satellites.

BIBLIOGRAPHY

[1] S. Aaronson. “Read the fine print”. *Nature Physics* **11** (2015), pp. 291–293.

[2] S. Aaronson. “Shadow tomography of quantum states”. (2017).
eprint: arXiv:1711.01053.

[3] S. Aaronson and D. Gottesman. “Identifying stabilizer states”. (2008).
eprint: pirsa.org/08080052.

[4] D. Aharonov, V. Jones, and Z. Landau. “A polynomial quantum algorithm for approximating the Jones polynomial”. *Algorithmica* **55** 3 (2009), pp. 395–421.

[5] J. Altepeter, E. Jeffrey, and P. Kwiat. “Photonic state tomography”. *Advances In Atomic, Molecular, and Optical Physics*. **52**. Academic Press, (2005), pp. 105–159.

[6] A. Ambainis. “Variable time amplitude amplification and a faster quantum algorithm for solving systems of linear equations”. *Proc. 29th Annual Symp. Theoretical Aspects of Computer Science*. (2012), pp. 636–647.

[7] G. G. Amosov, Y. A. Korennoy, and V. I. Man’ko. “Description and measurement of observables in the optical tomographic probability representation of quantum mechanics”. *Phys. Rev. A* **85** 052119 (2012).

[8] K. M. R. Audenaert et al. “Discriminating states: the quantum Chernoff bound”. *Phys. Rev. Lett.* **98** 160501 (2007).

[9] K. M. R. Audenaert et al. “Asymptotic error rates in quantum hypothesis testing”. *Comm. Math. Phys.* **279** 1 (2008), pp. 251–283.

[10] O. Axelsson and V. A. Barker. “Finite Element Solution of Boundary Value Problems: Theory and Computation”. Society for Industrial and Applied Mathematics, (2001).

[11] C. Bădescu, R. O’Donnell, and J. Wright. “Quantum state certification”. (2017).
eprint: arXiv:1708.06002.

[12] J.-D. Bancal et al. “Physical characterization of quantum devices from nonlocal correlations”. *Phys. Rev. A* **91** 022115 (2015).

[13] R. E. Bank and L. R. Scott. “On the conditioning of finite element equations with highly refined meshes”. *SIAM J. Numer. Anal.* **26** 6 (1989), pp. 1383–1394.

[14] S. Barnett and S. Croke. “Quantum state discrimination”. *Advances in Optics and Photonics* **1** 2 (2009), pp. 238–278.

[15] M. Bellini et al. “Towards higher precision and operational use of optical homodyne tomograms”. *Phys. Rev. A* **85** 052129 (2012).

[16] I. Bengtsson and H. Granström. “The frame potential, on average”. *Open Syst. Inform. Dynam.* **16** 02n03 (2009), pp. 145–156.

[17] C. Bennett et al. “Strengths and weaknesses of quantum computing”. *SIAM J. Comput.* **26** 5 (1997), pp. 1510–1523.

[18] M. Benzi, C. D. Meyer, and M. Tuma. “A sparse approximate inverse preconditioner for the conjugate gradient method”. *SIAM J. Sci. Comput.* **17** 5 (1996), pp. 1135–1149.

[19] M. Benzi and M. Tuma. “A comparative study of sparse approximate inverse preconditioners”. *App. Num. Math.* **30** 2-3 (1999), pp. 305–340.

[20] D. W. Berry, A. M. Childs, and R. Kothari. “Hamiltonian simulation with nearly optimal dependence on all parameters”. *Proc. 56th Annual Symp. Foundations of Computer Science*. (2015), pp. 792–809.

[21] D. W. Berry. “High-order quantum algorithm for solving linear differential equations”. *J. Phys. A: Math. Theor.* **47** 105301 (2014).

[22] B. Bertotti, L. Iess, and P. Tortora. “A test of general relativity using radio links with the Cassini spacecraft”. *Nature* **425** (2003), pp. 374–376.

[23] N. Birrell and P. Davies. “Quantum fields in curved space”. Cambridge University Press, (1984).

[24] A. Biswas et al. “LLCD operations using the Optical Communications Telescope Laboratory (OCTL)”. *Proc. SPIE*. **8971**. (2014).

[25] R. Blume-Kohout. “Optimal, reliable estimation of quantum states”. *New J. Phys.* **12** 043034 (2010).

[26] R. Blume-Kohout. “Robust error bars for quantum tomography”. (2012). eprint: `arXiv:1202.5270`.

[27] R. Blume-Kohout et al. “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography”. *Nature Comms.* **8** 14485 (2017).

[28] R. Blume-Kohout et al. “Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit”. (2013).
eprint: arXiv:1310.4492.

[29] P. Boyle, M. Broadie, and P. Glasserman. “Monte Carlo methods for security pricing”. *J. Econ. Dyn. Control* **21** 8–9 (1997), pp. 1267–1321.

[30] A. M. Branczyk et al. “Self-calibrating quantum tomography”. *New J. Phys.* **14** 085003 (2012).

[31] G. Brassard et al. “Quantum amplitude amplification and estimation”. *Quantum Computation and Quantum Information: A Millennium Volume* (2002), pp. 53–74.

[32] S. L. Braunstein et al. “Separability of very noisy mixed states and implications for NMR quantum computing”. *Phys. Rev. Lett.* **83** 5 (1999), pp. 1054–1057.

[33] S. L. Braunstein, A. Mann, and M. Revzen. “Maximal violation of Bell inequalities for mixed states”. *Phys. Rev. Lett.* **68** 22 (1992), pp. 3259–3261.

[34] S. C. Brenner and L. R. Scott. “The Mathematical Theory of Finite Element Methods”. Springer New York, (2008).

[35] A. Broadbent. “How to verify a quantum computation”. *Theory of Computation* **14** 11 (2018), pp. 1–37.

[36] A. Brodutch et al. “Post-Newtonian gravitational effects in optical interferometry”. *Phys. Rev. D* **91** 064041 (2015).

[37] D. Bruschi et al. “Spacetime effects on satellite-based quantum communications”. *Phys. Rev. D* **90** 045041 (2014).

[38] H. Buhrman et al. “Quantum fingerprinting”. *Phys. Rev. Lett.* **87** 167902 (2001).

[39] H. Buhrman et al. “Quantum property testing”. *SIAM J. Comput.* **37** 5 (2002), pp. 1387–1400.

[40] J. Calsamiglia and N. Lütkenhaus. “Maximum efficiency of a linear-optical Bell-state analyzer”. *App. Phys. B* **72** 1 (2001), pp. 67–71.

[41] Y. Cao et al. “Quantum algorithm and circuit design solving the Poisson equation”. *New J. Phys.* **15** 013021 (2013).

[42] J. Carolan et al. “On the experimental verification of quantum complexity in linear optics”. *Nature Photonics* **8** (2014), pp. 621–626.

[43] J. Carolan et al. “Universal linear optics”. *Science* **349** 6249 (2015), pp. 711–716.

[44] J. Cèa. “Approximation variationnelle des problèmes aux limites”. *Annales de l’Institut Fourier* **14** (1964), pp. 345–444.

[45] S. Chakraborty et al. “New results on quantum property testing”. *Proceedings of FSTTCS*. (2010), pp. 145–156.

[46] J. W. Chamberlain. “Physics of the aurora and airglow”. Academic Press, (1961).

[47] A. Chefles. “Quantum state discrimination”. *Contemporary Physics* **41** 6 (2001), pp. 401–424.

[48] L.-K. Chen et al. “Observation of ten-photon entanglement using thin BiB_3O_6 crystals”. *Optica* **4** 1 (2017), pp. 77–83.

[49] H. Chernoff. “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations”. *Ann. Math. Statist.* **23** 4 (1952), pp. 493–507.

[50] A. M. Childs, A. W. Harrow, and P. Wocjan. “Weak Fourier-Schur sampling, the hidden subgroup problem, and the quantum collision problem”. *Proc. 24th Symposium on Theoretical Aspects of Computer Science (STACS 2007)*. (2007), pp. 598–609.

[51] A. M. Childs and R. Kothari. “Limitations on the simulation of non-sparse Hamiltonians”. *Quantum Inf. Comput.* **10** (2009), pp. 669–684.

[52] A. M. Childs, R. Kothari, and R. D. Somma. “Quantum linear systems algorithm with exponentially improved dependence on precision”. *SIAM J. Comput.* **46** 6 (2017), pp. 1920–1950.

[53] M.-D. Choi. “Completely positive linear maps on complex matrices”. *Linear Algebra and its Applications* **10** 3 (1975), pp. 285–290.

[54] M. Christandl and R. Renner. “Reliable quantum state tomography”. *Phys. Rev. Lett.* **109** 120403 (2012).

[55] I. L. Chuang and M. A. Nielsen. “Prescription for experimental determination of the dynamics of a quantum black box”. *J. Mod. Opt.* **44** 11-12 (1997), pp. 2455–2467.

[56] P. Ciarlet. “The Finite Element Method for Elliptic Problems”. Elsevier, (1978).

[57] B. D. Clader, B. C. Jacobs, and C. R. Sprouse. “Preconditioned quantum linear system algorithm”. *Phys. Rev. Lett.* **110** 250504 (2013).

[58] E. M. Clarke, O. Grumberg, and D. Peleg. “Model Checking”. MIT Press, (1999).

[59] E. M. Clarke et al. “Progress on the state explosion problem in model checking”. *Informatics: 10 Years Back, 10 Years Ahead*. Springer Berlin Heidelberg, (2001), pp. 176–194.

[60] E. M. Clarke et al. “Model checking and the state explosion problem”. *Tools for Practical Software Verification (LASER 2011)*. Springer Berlin Heidelberg, (2012), pp. 1–30.

[61] J. Clauser et al. “Proposed experiment to test local hidden-variable theories”. *Phys. Rev. Lett.* **23** (1969), pp. 880–884.

[62] G. M. Clemence. “The relativity effect in planetary motions”. *Rev. Mod. Phys.* **19** 4 (1947), pp. 361–364.

[63] C. J. Clopper and E. S. Pearson. “The use of confidence or fiducial limits illustrated in the case of the binomial”. *Biometrika* **26** 4 (1934), pp. 404–413.

[64] A. Coladangelo. “A generalization of the CHSH inequality self-testing maximally entangled states of any local dimension”. (2018).
eprint: `arXiv:1803.05904`.

[65] A. Coladangelo, K. T. Goh, and V. Scarani. “All pure bipartite entangled states can be self-tested”. *Nature Comms.* **8** 15485 (2017).

[66] R. Colella, A. W. Overhauser, and S. A. Werner. “Observation of gravitationally induced quantum interference”. *Phys. Rev. Lett.* **34** 23 (1975), pp. 1472–1474.

[67] D. Collins et al. “Bell inequalities for arbitrarily high-dimensional systems”. *Phys. Rev. Lett.* **88** 040404 (2002).

[68] T. M. Cover and J. A. Thomas. “Elements of Information Theory”. Wiley-Interscience, (2006).

[69] M. Cramer et al. “Efficient quantum state tomography”. *Nature Comms.* **1** 149 (2010).

[70] A. W. Cross et al. “Scalable randomized benchmarking of non-Clifford gates”. *npj Quantum Information* **2** 16012 (2016).

[71] W. van Dam et al. “Self-testing of universal and fault-tolerant sets of quantum gates”. *Proc. 32nd Annual ACM Symp. Theory of Computing*. ACM, (2000), pp. 688–696.

[72] G. M. D’Ariano, C. Macchiavello, and N. Sterpi. “Systematic and statistical errors in homodyne measurements of the density matrix”. *J. Opt. B: Quantum Semiclassical Opt.* **9** 6 (1997).

[73] M. H. Devoret and R. J. Schoelkopf. “Superconducting circuits for quantum information: an outlook”. *Science* **339** 6124 (2013), pp. 1169–1174.

[74] H. Dittus and C. Lämmerzahl. “Experimental tests of the equivalence principle and Newton’s law in space”. *AIP Conf. Proc.* **758** 1 (2005), pp. 95–112.

[75] V. V. Dobrovitski et al. “Bootstrap tomography of the pulses for quantum control”. *Phys. Rev. Lett.* **105** 077601 (2010).

[76] J. Emerson et al. “Symmetrized characterization of noisy quantum processes”. *Science* **317** 5846 (2007), pp. 1893–1896.

[77] A. Ern and J.-L. Guermond. “Theory and Practice of Finite Elements”. Springer, (2013).

[78] F. Ewert and P. van Loock. “3/4-efficient Bell measurement with passive linear optics and unentangled ancillae”. *Phys. Rev. Lett.* **113** 140403 (2014).

[79] P. Faist and R. Renner. “Practical and reliable error bars in quantum tomography”. *Physical Review Letters* **117** 010404 (2016).

[80] T. Ferreiro, J. Sun, and D. T. Reid. “Frequency stability of a femtosecond optical parametric oscillator frequency comb”. *Optics Express* **19** 24 (2011), pp. 24159–25164.

[81] C. Ferrie. “High posterior density ellipsoids for quantum states”. *New J. Phys.* **16** 023006 (2014).

[82] C. Ferrie and R. Blume-Kohout. “Minimax quantum tomography: estimators and relative entropy bounds”. *Phys. Rev. Lett.* **116** 090407 (2016).

[83] J. F. Fitzsimons, M. Hajdušek, and T. Morimae. “Post-hoc verification of quantum computation”. *Phys. Rev. Lett.* **120** 040501 (2018).

[84] S. T. Flammia and Y.-K. Liu. “Direct fidelity estimation from few Pauli measurements”. *Phys. Rev. Lett.* **106** 230501 (2011).

[85] S. T. Flammia et al. “Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators”. *New J. Phys.* **14** 095022 (2012).

[86] A. Gheorghiu, T. Kapourniotis, and E. Kashefi. “Verification of quantum computation: an overview of existing approaches”. *Theory Comput. Syst.* (2018), pp. 1–94.

[87] A. Gheorghiu, P. Wallden, and E. Kashefi. “Rigidity of quantum steering and one-sided device-independent verifiable quantum computation”. *New J. Phys.* **19** 023043 (2017).

[88] E. Gibney. “Chinese satellite is one giant step for the quantum internet”. *Nature* **535** (2016), pp. 478–479.

[89] D. Gottesman. “Class of quantum error-correcting codes saturating the quantum Hamming bound”. *Phys. Rev. A* **54** 3 (1996), pp. 1862–1868.

[90] D. Gottesman. “Stabilizer Codes and Quantum Error Correction”. PhD thesis. Caltech, 1997.

[91] C. Granade, J. Combes, and D. G. Cory. “Practical Bayesian tomography”. *New J. Phys.* **18** 033024 (2016).

[92] C. Granade, C. Ferrie, and S. T. Flammia. “Practical adaptive quantum tomography”. *New J. Phys.* **19** 113017 (2017).

[93] C. Granade et al. “QInfer: Statistical inference software for quantum applications”. *Quantum* **1** (2016).

[94] D. Greenbaum. “Introduction to gate set tomography”. (2015).
eprint: arXiv:1509.02921.

[95] D. Gross, S. Nezami, and M. Walter. “Schur-Weyl duality for the Clifford group wih applications: property testing, a robust Hudson theorem and de Finetti representations”. (2017).
eprint: arXiv:1712.08628.

[96] D. Gross et al. “Quantum state tomography via compressed sensing”. *Phys. Rev. Lett.* **105** 150401 (2010).

[97] L. Grover and J. Radhakrishnan. “Is partial quantum search of a database any easier?” *Proc. 17th Annual ACM Symp. Parallelism in Algorithms and Architectures*. ACM, (2005), pp. 186–194.

[98] L. Grover and T. Rudolph. “Creating superpositions that correspond to efficiently integrable probability distributions”. (2002).
eprint: arXiv:0208112.

[99] B. Guo and I. Babuška. “The h-p version of the finite element method”. *Computational Mechanics* **1** 3 (1986), pp. 203–220.

[100] J. Haah et al. “Sample-optimal tomography of quantum states”. *Proc. 48th Annual ACM Symp. Theory of Computing*. ACM, (2016), pp. 913–925.

[101] R. Haber et al. “A general two-dimensional, graphical finite element preprocessor utilizing discrete transfinite mappings”. *Int. J. Numer. Meth. Eng.* **17** 7 (1981), pp. 1015–1044.

[102] H. Häffner et al. “Scalable multiparticle entanglement of trapped ions”. *Nature* **438** 7068 (2005), pp. 643–646.

[103] M. A. Harrison. “Introduction to Switching and Automata Theory”. McGraw-Hill, (1965).

[104] A. W. Harrow, A. Hassidim, and S. Lloyd. “Quantum algorithm for linear systems of equations”. *Phys. Rev. Lett.* **15** 150502 (2009).

[105] A. Harrow, C. Lin, and A. Montanaro. “Sequential measurements, disturbance and property testing”. *Proc. 28th ACM-SIAM Symp. Discrete Algorithms*. (2017), pp. 1598–1611.

[106] A. Harrow and A. Montanaro. “Testing product states, quantum Merlin-Arthur games and tensor optimization”. *J. ACM* **60** 1 (2013).

[107] S. W. Hawking. “Particle creation by black holes”. *Comm. Math. Phys.* **43** 3 (1975), pp. 199–220.

[108] M. Hayashi. “Asymptotic estimation theory for a finite-dimensional pure state model”. *J. Phys. A: Math. Gen.* **31** 20 (1998), pp. 4633–4655.

[109] M. Hayashi and T. Morimae. “Verifiable measurement-only blind quantum computing with stabilizer testing”. *Phys. Rev. Lett.* **115** 220502 (2015).

[110] M. Hein. “Entanglement in graph states”. PhD thesis. University of Innsbruck, 2005.

[111] J. Helsen et al. “Multi-qubit randomised benchmarking using few samples”. (2017).
eprint: `arXiv:1701.04299`.

[112] C. W. Helstrom. “Quantum Detection and Estimation Theory”. Academic Press, New York, (1976).

[113] F. Hiai and D. Petz. “The proper formula for relative entropy and its asymptotics in quantum probability”. *Comm. Math. Phys.* **143** 1 (1991), pp. 99–114.

[114] W. Hoeffding. “Probability inequalities for sums of bounded random variables”. *J. Am. Stat. Assoc.* **58** 301 (1963), pp. 13–30.

[115] H. F. Hofmann. “Complementary classical fidelities as an efficient criterion for the evaluation of experimentally realized quantum operations”. *Phys. Rev. Lett.* **94** 160504 (2005).

[116] A. S. Holevo. “Bounds for the quantity of information transmitted by a quantum communication channel”. *Problems Inform. Transmission* **9** (1973), pp. 177–183.

[117] Z. Hradil. “Quantum-state estimation”. *Phys. Rev. A* **55** 3 (1997), R1561–R1564.

[118] F. Huszár and N. M. T. Houlsby. “Adaptive Bayesian quantum tomography”. *Phys. Rev. A* **85** 052120 (2012).

[119] C. J. Isham. “Canonical quantum gravity and the problem of time”. *Integrable Systems, Quantum Groups, and Quantum Field Theories*. **409**. Springer Netherlands, (1993), pp. 157–287.

[120] A. Jamiołkowski. “Linear transformations which preserve trace and positive semidefiniteness of operators”. *Rep. Math. Phys.* **3** 4 (1972), pp. 275–278.

[121] J. Jiang et al. “Frequency measurement of acetylene-stabilized lasers using a femtosecond optical comb without carrier-envelope offset frequency control”. *Optics Express* **13** 6 (2005), pp. 1958–1965.

[122] L. Jiang and C. Li. “Mathematical Modeling and Methods of Option Pricing”. World Scientific, (2005).

[123] S. Jordan. “Quantum algorithm zoo”. (2018).
eprint: `https://math.nist.gov/quantum/zoo/`.

[124] J. Kaniewski. “Analytic and (nearly) optimal self-testing bounds for the CHSH and Mermin inequalities”. *Phys. Rev. Lett.* **117** 070402 (2016).

[125] P. Kaye and M. Mosca. “Quantum networks for generating arbitrary quantum states”. *Optical Fiber Communication Conference and International Conference on Quantum Information*. OSA, (2001).

[126] J. Kelly. “Engineering superconducting qubit arrays for quantum supremacy”. *APS March meeting 2018*. APS, (2018).

[127] I. Kerenidis and A. Prakash. “Quantum recommendation systems”. (2016).
eprint: `arXiv:1603.08675`.

[128] I. Kerenidis and A. Prakash. “Quantum gradient descent for linear systems and least squares”. (2017).
eprint: `arXiv:1704.04992`.

[129] E. Knill et al. “Randomized benchmarking of quantum gates”. *Phys. Rev. A* **77** 012307 (2008).

[130] J. Kouba. “Improved relativistic transformations in GPS”. *GPS Solutions* **8** 3 (2004), pp. 170–180.

[131] R. Kueng, H. Rauhut, and U. Terstiege. “Low rank matrix recovery from rank one measurements”. *App. Comput. Harmon. Anal.* **42** 1 (2017), pp. 88–116.

[132] A. Laing and J. L. O’Brien. “Super-stable tomography of any linear optical device” (2012).
eprint: `arXiv:1208.2868`.

[133] A. Laing et al. “High-fidelity operation of quantum photonic circuits”. *App. Phys. Lett.* **97** 211109 (2010).

[134] L. LaPiana and F. Bauer. *Mars Climate Orbiter Mishap Investigation Board: Phase I report*. Tech. rep. NASA, 1999.

[135] P. D. Lax and A. N. Milgram. “Parabolic equations”. *Ann. Math. Stud.* **33** (1954), pp. 167–190.

[136] P. L’Ecuyer. “Quasi-Monte Carlo methods with applications in finance”. *Finance and Stochastics* **13** 3 (2009), pp. 307–349.

[137] S.-H. Lee et al. “Active compensation of large dispersion of femtosecond pulses for precision laser ranging”. *Optics Express* **19** 5 (2011), pp. 4002–4008.

[138] C. Leinert et al. “The 1997 reference of diffuse night sky brightness”. *Astron. Astrophys. Suppl. Ser.* **127** (1998), pp. 1–99.

[139] U. Leonhardt et al. “Sampling of photon statistics and density matrix using homodyne detection”. *Optics Comms.* **127** 1-3 (1996), pp. 144–160.

[140] S. Leyton and T. Osborne. “A quantum algorithm to solve nonlinear differential equations”. (2008).
eprint: arXiv:0812.4423.

[141] S. Li, P. Rui, and R. Chen. “An effective sparse approximate inverse preconditioner for vector finite element analysis of 3D EM problems”. *Antennas and Propagation Society International Symp.* IEEE, (2006), pp. 1765–1768.

[142] S. Lloyd, M. Mohseni, and P. Rebentrost. “Quantum algorithms for supervised and unsupervised machine learning”. (2013).
eprint: arXiv:1307.0411.

[143] M. Er-long et al. “Background noise of satellite-to-ground quantum key distribution”. *New. J. Phys.* **7** 215 (2005).

[144] R. Loudon. “The Quantum Theory of Light”. Oxford University Press, (2003).

[145] A. I. Lvovsky and M. G. Raymer. “Continuous-variable optical quantum-state tomography”. *Rev. Mod. Phys.* **81** 1 (2009), pp. 299–332.

[146] E. Magesan, J. M. Gambetta, and J. Emerson. “Scalable and robust randomized benchmarking of quantum processes”. *Phys. Rev. Lett.* **106** 180504 (2011).

[147] E. Magesan, J. M. Gambetta, and J. Emerson. “Characterizing quantum gates via randomized benchmarking”. *Phys. Rev. A* **85** 042311 (2012).

[148] F. Magniez et al. “Self-testing of quantum circuits”. *Automata, Languages and Programming (ICALP 2006)*. **4051**. Springer, (2006), pp. 72–83.

[149] D. H. Mahler et al. “Adaptive quantum state tomography improves accuracy quadratically”. *Phys. Rev. Lett.* **111** 183601 (2013).

[150] I. Malitson. “Interspecimen comparison of the refractive index of fused silica”. *J. Opt. Soc. Am.* **55** 10 (1965), pp. 1205–1209.

[151] A. Mantri et al. “Flow ambiguity: a path towards classically driven blind quantum computation”. *Phys. Rev. X* **7** 031004 (2017).

[152] F. Marsili et al. “Detecting single infrared photons with 93% system efficiency”. *Nature Photonics* **7** (2013), pp. 210–214.

[153] K. Mayer and E. Knill. “Quantum process fidelity bounds from sets of input states”. (2018).
eprint: arXiv:1805.04491.

[154] D. Mayers and A. Yao. “Self testing quantum apparatus”. *Quantum Inf. Comput.* **4** 4 (2004), pp. 273–286.

[155] M. McKague. “Self-testing graph states”. *Proc. 6^{th} Conference on the Theory of Quantum Computation, Communication, and Cryptography*. Springer Berlin, (2010), pp. 104–120.

[156] M. McKague, T. H. Yang, and V. Scarani. “Robust self-testing of the singlet”. *J. Phys. A: Math. Theor.* **45** 455304 (2012).

[157] M. McKague. “Interactive proofs for BQP via self-tested graph states”. *Theory of Computation* **12** 3 (2016), pp. 1–42.

[158] M. McKague. “Self-testing in parallel”. *New J. Phys.* **18** 045013 (2016).

[159] S. T. Merkel et al. “Self-consistent quantum process tomography”. *Phys. Rev. A* **87** 062119 (2013).

[160] F. Mintert, M. Kuś, and A. Buchleitner. “Concurrence of mixed multipartite quantum states”. *Phys. Rev. Lett.* **95** 260502 (2005).

[161] A. Montanaro. “Quantum algorithms: an overview”. *npj Quantum Information* **2** 15023 (2016).

[162] A. Montanaro. “Learning stabilizer states by Bell sampling”. (2017).
eprint: arXiv:1707.04012.

[163] A. Montanaro and R. de Wolf. “A survey of quantum property testing”. *Theory of Computing Library Graduate Surveys* **7** (2016), pp. 1–81.

[164] A. Montanaro and S. Pallister. “Quantum algorithms and the finite element method”. *Phys. Rev. A* **93** 032324 (2016).

[165] A. Montanaro and D. J. Shepherd. “Hadamard gates and amplitudes of computational basis states”. (2006).
eprint: https://www.scottaaronson.com/hadamard.pdf.

[166] P. Mosley et al. “Heralded generation of ultrafast single photons in pure quantum states”. *Phys. Rev. Lett.* **100** 133601 (2008).

[167] A. Natarajan and T. Vidick. “A quantum linearity test for robustly verifying entanglement”. *Proc. 49th Annual ACM Symp. Theory of Computing*. ACM, (2017), pp. 1003–1015.

[168] C. Neill et al. “A blueprint for demonstrating quantum supremacy with superconducting qubits”. *Science* **360** 6385 (2018), pp. 195–199.

[169] E. Nielsen et al. “Gate set tomography on more than two qubits”. *APS March meeting 2018*. APS, (2018).

[170] M. A. Nielsen and I. L. Chuang. “Quantum Computation and Quantum Information: 10th Anniversary Edition”. Cambridge University Press, (2010).

[171] M. Nussbaum and A. Szkoła. “The Chernoff lower bound for symmetric quantum hypothesis testing”. *Ann. Statist.* **37** 2 (2009), pp. 1040–1057.

[172] B. Odom et al. “New measurement of the electron magnetic moment using a one-electron quantum cyclotron”. *Phys. Rev. Lett.* **97** 030801 (2006).

[173] R. O’Donnell and J. Wright. “Quantum spectrum testing”. *Proc. 47th Annual ACM Symp. Theory of Computing*. ACM, (2015), pp. 529–538.

[174] R. O’Donnell and J. Wright. “Efficient quantum tomography”. *Proc. 48th Annual ACM Symp. Theory of Computing*. ACM, (2016), pp. 899–912.

[175] R. O’Donnell and J. Wright. “Efficient quantum tomography II”. *Proc. 49th Annual ACM Symp. Theory of Computing*. ACM, (2017), pp. 962–974.

[176] T. Ogawa and H. Nagaoka. “Strong converse and Stein’s lemma in quantum hypothesis testing”. *IEEE Trans. Inf. Theory* **46** 7 (2000), pp. 2428–2433.

[177] K. F. Pál, T. Vértesi, and M. Navascués. “Device-independent tomography of multipartite quantum states”. *Phys. Rev. A* **90** 042340 (2014).

[178] S. Pallister, N. Linden, and A. Montanaro. “Optimal verification of entangled states with local measurements”. *Phys. Rev. Lett.* **120** 170502 (2018).

[179] S. Pallister et al. “A blueprint for a simultaneous test of quantum mechanics and general relativity in a space-based quantum optics experiment”. *EPJ Quantum Technology* **4** 2 (2017).

[180] M. Pelusi et al. “Fourth-order dispersion compensation for 250-fs pulse transmission over 139-km optical fiber”. *IEEE Photon. Tech. Lett.* **12** 7 (2000), pp. 795–797.

[181] A. Phillips et al. “The near-infrared sky emission at the South Pole in winter”. *Astrophys. J.* **527** (1999), pp. 1009–1022.

[182] X. W. Ping and T.-J. Cui. “The factorized sparse approximate inverse preconditioned conjugate gradient algorithm for finite element analysis of scattering problems”. *Prog. Electromagn. Res.* **98** (2009), pp. 15–31.

[183] S. Popescu and D. Rohrlich. “Which states violate Bell’s inequality maximally?” *Phys. Lett. A* **169** 6 (1992), pp. 411–414.

[184] R. Pound and G. Rebka Jr. “Apparent weight of photons”. *Phys. Rev. Lett.* **4** 337 (1960).

[185] J. F. Poyatos, J. I. Cirac, and P. Zoller. “Complete characterization of a quantum process: the two-bit quantum gate”. *Phys. Rev. Lett.* **78** 2 (1997), pp. 390–393.

[186] A. Prakash. “Quantum Algorithms for Linear Algebra and Machine Learning”. PhD thesis. University of California, Berkeley, 2014.

[187] T. Proctor et al. “What randomized benchmarking actually measures”. *Phys. Rev. Lett.* **119** 130502 (2017).

[188] T. Proctor et al. “Direct randomised benchmarking for multi-qubit devices”. (2018).
eprint: `arXiv:1807.07975`.

[189] Project Management Institute. “A guide to the project management book of knowledge (PMBOK guide)”. Project Management Institute, (2004).

[190] X. Qiang et al. “Large-scale silicon quantum photonics implementing arbitrary two-qubit processing”. (2018, in press).

[191] S. S. Rao. “The Finite Element Method in Engineering”. Butterworth-Heinemann, (2005).

[192] D. M. Reich, G. Gualdi, and C. P. Koch. “Minimum number of input states required for quantum gate characterization”. *Phys. Rev. A* **88** 042309 (2013).

[193] B. W. Reichardt, F. Unger, and U. Vazirani. “Classical command of quantum systems”. *Nature* **496** (2013), pp. 456–460.

[194] D. Rosset et al. “Imperfect measurement settings: implications for quantum state tomography and entanglement witnesses”. *Phys. Rev. A* **86** 062325 (2012).

[195] D. Sadot et al. “Restoration of thermal images distorted by the atmosphere, using predicted atmospheric modulation transfer function”. *Infrared Phys. Technol.* **36** 2 (1995), pp. 565–576.

[196] A. Salavrakos et al. “Bell inequalities tailored to maximally entangled states”. *Phys. Rev. Lett.* **119** 040402 (2017).

[197] R. Santagati et al. “Witnessing eigenstates for quantum simulation of Hamiltonian spectra”. *Science Advances* **4** 1 (2018).

[198] R. Santagati et al. “Silicon photonic processor of two-qubit entangling quantum logic”. *J. Opt.* **19** 114006 (2017).

[199] A. Scherer et al. “Concrete resource analysis of the quantum linear-system algorithm used to compute the electromagnetic scattering cross section of a 2D target”. *Quantum Inf. Process.* **16** 60 (2017).

[200] T. Schmitt-Manderbach. “Long distance free-space quantum key distribution”. PhD thesis. LMU Munich, 2007.

[201] C. Schwemmer et al. “Systematic errors in current quantum state tomography tools”. *Phys. Rev. Lett.* **114** 080403 (2015).

[202] J. Shang, Z. Zhang, and H. K. Ng. “Superfast maximum likelihood reconstruction for quantum tomography”. *Phys. Rev. A* **95** 062336 (2017).

[203] J. A. Shaw. “Radiometry and the Friis transmission equation”. *Am. J. Phys.* **81** 1 (2013), pp. 33–37.

[204] L. Shen et al. “Design and optimization of photonic crystal fibers for broad-band dispersion compensation”. *IEEE Photon. Tech. Lett.* **15** 4 (2003), pp. 540–542.

[205] J. Shewchuk. *An Introduction to the Conjugate Gradient Method without the Agonizing Pain*. Tech. rep. CMU-CS-TR-94-125. Carnegie Mellon University, 1994.

[206] M. P. da Silva, O. Landon-Cardinal, and D. Poulin. “Practical characterization of quantum devices without tomography”. *Phys. Rev. Lett.* **107** 210404 (2011).

[207] J. W. Silverstone et al. “On-chip quantum interference between silicon photon-pair sources”. *Nature Photonics* **8** (2013), pp. 104–108.

[208] J. A. Smolin, J. M. Gambetta, and G. Smith. “Efficient method for computing the maximum-likelihood quantum state from measurements with additive Gaussian noise”. *Phys. Rev. Lett.* **108** 070502 (2012).

[209] S. Straupe. “Adaptive quantum tomography”. *JETP Lett.* **104** 7 (2016), pp. 510–522.

[210] G. I. Struchalin et al. “Experimental adaptive quantum tomography of two-qubit states”. *Phys. Rev. A* **93** 012103 (2016).

[211] Z.-E. Su et al. “Multiphoton interference in quantum Fourier transform circuits and applications to quantum metrology”. *Phys. Rev. Lett.* **119** 080502 (2017).

[212] T. Sugiyama. “Finite Sample Analysis in Quantum Estimation”. Springer, (2014).

[213] T. Sugiyama, P. S. Turner, and M. Murao. “Precision-guaranteed quantum tomography”. *Phys. Rev. Lett.* **111** 160406 (2013).

[214] I. Šupić et al. “A simple approach to self-testing multipartite entangled states”. (2017).
eprint: `arXiv:1707.06534`.

[215] Y. M. Timofeyev and A. V. Vasilev. “Theoretical Fundamentals of Atmospheric Optics”. Cambridge International Science Publishing, (2008).

[216] G. Tologlou et al. “Distortion-less 610 fs pulse transmission over 160 km SSMF-DCF using wavelength selective switch for compensation of chromatic dispersion”. *IEEE Photonics Conf.* IEEE, (2011), pp. 829–830.

[217] G. Tóth and O. Gühne. “Detecting genuine multipartite entanglement with two local measurements”. *Phys. Rev. Lett.* **94** 060501 (2005).

[218] G. Tóth and O. Gühne. “Entanglement detection in the stabilizer formalism”. *Phys. Rev. A* **72** 022340 (2005).

[219] G. Tóth et al. “Permutationally invariant quantum tomography”. *Phys. Rev. Lett.* **105** 250403 (2010).

[220] B. S. Tsirelson. “Quantum generalizations of Bell’s inequality”. *Lett. Math. Phys.* **4** 2 (1980), pp. 93–100.

[221] M. van Uffelen et al. “Feasibility study for distributed dose monitoring in ionizing radiation environments with standard and custom-made optical fibers”. *Proc. SPIE: Photonics for Space Environments*. **4823**. (2002).

[222] L. Vaidman and N. Yoran. “Methods for reliable teleportation”. *Phys. Rev. A* **59** 1 (1999), pp. 116–125.

[223] D. Wallace and R. Fujii. “Software verification and validation: an overview”. *IEEE Software* **6** 3 (1989), pp. 10–17.

[224] G. Wang. “Property testing of unitary operators”. *Phys. Rev. A* **84** 052328 (2011).

[225] G. Wang. “Efficient quantum algorithms for analyzing large sparse electrical networks”. *Quantum Inf. Comput.* **17** 11-12 (2017).

[226] J. Wang et al. “Gallium arsenide (GaAs) quantum photonic waveguide circuits”. *Opt. Comms.* **327** (2014), pp. 49–55.

[227] J. Wang et al. “Multidimensional quantum entanglement with large-scale integrated optics”. *Science* **360** 6386 (2018), pp. 285–291.

[228] X.-L. Wang et al. “Experimental ten-photon entanglement”. *Phys. Rev. Lett.* **117** 210502 (2016).

[229] X.-L. Wang et al. “18-qubit entanglement with six photons’ three degrees of freedom”. *Phys. Rev. Lett.* **120** 260502 (2018).

[230] Y. Wang, X. Wu, and V. Scarani. “All the self-testings of the singlet for two binary measurements”. *New J. Phys.* **18** 025021 (2016).

[231] C. M. Will. “The confrontation between general relativity and experiment”. *Living Rev. Relativ.* **17** 4 (2014).

[232] L. Wossnig, Z. Zhao, and A. Prakash. “Quantum linear system algorithm for dense matrices”. *Phys. Rev. Lett.* **120** 050502 (2018).

[233] X. Wu et al. “Robust self-testing of the three-qubit W state”. *Phys. Rev. A* **90** 042339 (2014).

[234] T. H. Yang and M. Navascués. “Robust self-testing of unknown quantum systems into any entangled two-qubit states”. *Phys. Rev. A* **87** 050102 (2013).

[235] T. H. Yang et al. “Robust and versatile black-box certification of quantum devices”. *Phys. Rev. Lett.* **113** 040401 (2014).

[236] C. Zalka. “Simulating quantum systems on a quantum computer”. *Proc. Royal Soc. A* **454** 1969 (1998), pp. 313–322.

[237] M. Zych et al. “General relativistic effects in quantum interference of photons”. *Class. Quantum Grav.* **29** 224010 (2012).

APPENDIX A

VERIFICATION OF QUANTUM STATES: ADDITIONAL CALCULATIONS

A.1 Bell tests

We first prove Theorem 2 in § 2.1.3 relating the copy complexity for a CHSH test with the mean squared error (MSE).

Theorem 2 (restated) (Bell test MSE [212])**.** *Given a CHSH game carried out on n identical copies of a two-qubit state with expected Bell parameter S_{CHSH}, to bound the MSE by Δ^2 it is necessary for the verifier to measure a number of copies*

$$n \geq \frac{16 - |S_{CHSH}|^2}{\Delta^2}. \tag{A.1}$$

Proof. Suppose that we have a sequence of random variables, $\{X_1 \ldots X_n\}$, such that their expectation values are $\mathbb{E}(X_i) = \mu_i$. Then define the covariance matrix Γ, with elements given by

$$\Gamma_{jk} = \mathbb{E}[(X_j - \mu_j)(X_k - \mu_k)]. \tag{A.2}$$

We would like to take the outcomes from the sequence of random variables, and combine them in such a way as to get an estimate of a single quantity. In particular, using the vectorial shorthand $\mathbf{X} = (X_1, X_2, \ldots X_n)^\top$ and $\mu = (\mu_1, \mu_2, \ldots \mu_n)^\top$, we will be interested in the weighted average $\mathbf{a} \cdot \mathbf{X}$, for some vector of positive weights $\mathbf{a}$. We assess the quality of the estimate using the mean squared error (MSE), denoted Δ^2;

which in this instance is given by

$$\begin{aligned}\Delta^2 &:= \mathbb{E}[|\mathbf{a}\cdot\mathbf{X}-\mathbf{a}\cdot\mu|^2] \qquad \text{(A.3)}\\ &= \mathbb{E}\left[\sum_{j,k} a_j a_k X_j X_k - a_j a_k \mu_j X_k - a_j a_k \mu_k X_j + a_j a_k \mu_j \mu_k\right]\\ &= \mathbb{E}\left[\sum_{j,k} a_j (X_j-\mu_j)(X_k-\mu_k) a_k\right]\\ &= \frac{1}{n}\sum_{j,k} a_j \mathbb{E}[(X_j-\mu_j)(X_k-\mu_k)] a_k\\ &= \frac{\mathbf{a}^\top \Gamma \mathbf{a}}{n}. \qquad \text{(A.4)}\end{aligned}$$

In a CHSH test, each trial is a random variable drawn from a distribution that is dependent on (i) Alice and Bob's choice of measurement setting; and (ii) measurement outcomes for that particular setting. Index the former with $i \in \{0,1\}$ and $j \in \{0,1\}$ for Alice and Bob, respectively, and the latter with outcomes a and b. If we assume that each choice of setting for Alice and Bob is equally likely, then write the unconditional probability distribution governing outcomes as $p(i,j,a,b) = \frac{1}{4}p(a,b|i,j)$. Then the CHSH parameter is

$$S_{CHSH} = \sum_{i,j}\sum_{a,b} 4p(i,j,a,b)\cdot a\cdot b\cdot \delta(i,j), \qquad \text{(A.5)}$$

where the quantity $\delta(i,j) = -1$ for $i=1, j=1$ and $\delta(i,j)=1$ otherwise. Denote the random variable governing the drawing of a sample from the distribution $p(i,j,a,b)$ as $X(i,j,a,b)$; then the experimental estimate of the CHSH parameter is

$$S^{est}_{CHSH} = \sum_{i,j}\sum_{a,b} 4X(i,j,a,b)\cdot a\cdot b\cdot \delta(i,j). \qquad \text{(A.6)}$$

Then from Eq. A.4, the mean squared error between the true and estimated parameter is given by

$$\Delta^2_{CHSH} = \mathbb{E}[(S_{CHSH} - S^{est}_{CHSH})^2] = \frac{\mathbf{c}^\top \Gamma(p)\mathbf{c}}{n}, \qquad \text{(A.7)}$$

where $\mathbf{c}$ is a vector with elements $c_{ijab} = \delta(i,j)\cdot a\cdot b$. Using the identity for the covariance matrix that $\Gamma_{kl} = \mathbb{E}[(X_k-\mu_k)(X_l-\mu_l)] = \mathbb{E}[X_k X_l] - \mu_k\mu_l$, we can rewrite this expression as

$$\begin{aligned}\Delta^2_{CHSH} = \mathbb{E}[|S_{CHSH} - S^{est}_{CHSH}|^2] = \frac{\mathbf{c}^\top\Gamma(p)\mathbf{c}}{n} &= \frac{1}{n}\sum_{i,j,a,b}\left(|c_{ijab}|^2 - |c^2_{ijab}p(i,j,a,b)|^2\right)\\ &= \frac{16-|S_{CHSH}|^2}{n}. \qquad \text{(A.8)}\end{aligned}$$

Hence, if we wish to carry out a CHSH test such that the mean squared error is at most Δ^2, then we need a number of trials

$$n \geq \frac{16 - |S_{CHSH}|^2}{\Delta^2}. \tag{A.9}$$

□

We now prove Lemma 3, and convert the MSE bound into a fidelity bound.

Lemma 3 (restated). *Consider a CHSH test carried out on n copies of an output state claimed to be $|\psi_\theta\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, where we must use this test in order to discriminate from a state σ such that $F(|\psi_\theta\rangle, \sigma) = 1 - \epsilon$, $0 \leq \epsilon \leq \frac{1}{2}$. Then the root MSE between these two cases is given by*

$$\Delta = |\epsilon(S_\theta - S_\phi) - \sqrt{\epsilon(1-\epsilon)}S_j| \tag{A.10}$$

for ϵ-independent parameters S_θ, S_ϕ and S_j that depend upon $|\psi_\theta\rangle$, σ, and the choice of calibration for the CHSH test. Thus there is a lower bound on the copy complexity:

$$n \geq \frac{16 - S_\theta^2}{\epsilon^2(S_\theta - S_\phi - S_j)^2} \tag{A.11}$$

copies are necessary to verify that σ is within infidelity ϵ to the target state $|\psi_\theta\rangle$ using a CHSH test.

Proof. We can decompose a generic σ into the following form:

$$\sigma = (1-\epsilon)|\psi_\theta\rangle\langle\psi_\theta| + \epsilon\sum_{j,k} c_{jk}|\phi_j\rangle\langle\phi_k| + \sqrt{\epsilon(1-\epsilon)}\sum_k (b_k|\psi_\theta\rangle\langle\phi_k| + b_k^*|\phi_k\rangle\langle\psi_\theta|). \tag{A.12}$$

Then if we denote the CHSH operator by S, by linearity of expectation we have that

$$\begin{aligned}\operatorname{tr}(S\sigma) &= (1-\epsilon)\operatorname{tr}(S|\psi_\theta\rangle\langle\psi_\theta|) + \epsilon\sum_{j,k} c_{jk}\operatorname{tr}(S|\phi_j\rangle\langle\phi_k|) \\ &\quad + \sqrt{\epsilon(1-\epsilon)}\sum_k (b_k\operatorname{tr}(S|\psi_\theta\rangle\langle\phi_k|) + b_k^*\operatorname{tr}(S|\phi_k\rangle\langle\psi_\theta|)) \\ &:= (1-\epsilon)S_\theta + \epsilon S_\phi + \sqrt{\epsilon(1-\epsilon}S_j. \end{aligned} \tag{A.13}$$

Then the root MSE between $|\psi_\theta\rangle$ and σ is given by

$$\Delta = |S_\theta - (1-\epsilon)S_\theta - \epsilon S_\phi - \sqrt{\epsilon(1-\epsilon}S_j| = |\epsilon(S_\theta - S_\phi) - \sqrt{\epsilon(1-\epsilon)}S_j|. \tag{A.14}$$

Using Theorem 2, the copy complexity can then be bounded by

$$n \geq \frac{16 - S_\theta^2}{\Delta^2} = \frac{16 - S_\theta^2}{\epsilon(\sqrt{\epsilon}(S_\theta - S_\phi) - \sqrt{1-\epsilon}S_j)^2}. \tag{A.15}$$

For $0 \leq \epsilon \leq \frac{1}{2}$, we have that $\sqrt{\epsilon} \leq \sqrt{1-\epsilon}$, and so we can give the tidier lower bound

$$n \geq \frac{16 - S_\theta^2}{\epsilon^2(S_\theta - S_\phi - S_j)^2}. \tag{A.16}$$

□

A.2 Direct fidelity estimation

We now give a proof of Theorem 4 in § 2.1.4, that there is a sufficient choice of parameters for the verifier in the DFE protocol that guarantee that the estimated fidelity is close to the true fidelity of the output state with the target, with high probability.

Theorem 4 (restated) (Direct fidelity estimation [84])**.** *Suppose the verifier makes the following choices for the parameters ℓ, m_i, C_i and p_i:*

$$\ell = \frac{8}{\epsilon^2\delta}; \quad p_i = \frac{|\langle\psi|\sigma_i|\psi\rangle|^2}{d}; \quad m_i = \frac{d\delta}{4|\langle\psi|\sigma_i|\psi\rangle|^2}\log\frac{4}{\delta}; \quad C_i = \frac{1}{\langle\psi|\sigma_i|\psi\rangle}. \tag{A.17}$$

Then it is guaranteed that

$$Pr[|\tilde{X} - F| \geq \epsilon] \leq \delta. \tag{A.18}$$

Proof. Firstly, we show that the estimate is close to the true fidelity in expectation; then we show that an imperfect estimate is sufficiently close to its expectation value. The expectation of the quantity X_i is given by

$$\mathbb{E}(X_i) = \frac{C_i}{m_i}\sum_j \mathbb{E}[a_{ij}] = \frac{C_i}{m_i}\sum_j \mathrm{tr}(\rho\sigma_i) = \frac{C_i}{m_i} m_i \,\mathrm{tr}(\rho\sigma_i) = \frac{\mathrm{tr}(\rho\sigma_i)}{\langle\psi|\sigma_i|\psi\rangle}, \tag{A.19}$$

and so the expectation of X distributed over p is

$$\mathbb{E}_p(X) = \sum_i p_i \mathbb{E}(X_i) = \sum_i \frac{\langle\psi|\sigma_i|\psi\rangle\,\mathrm{tr}(\rho\sigma_i)}{d} = \mathrm{tr}(\rho|\psi\rangle\langle\psi|) = F. \tag{A.20}$$

Suppose now that all the X_i's have been estimated perfectly, but we only have a finite number ℓ of them; then call the overall estimate X. We wish to bound the difference between X and F. So, use Chebyshev's inequality; that is, for a random variable R and parameter $\omega > 0$,

$$\Pr[|R - \mathbb{E}(R)| \geq \omega \mathrm{Var}(R)] \leq \frac{1}{\omega^2}. \tag{A.21}$$

We would like to show

$$\Pr\left[|X - F| \geq \frac{\epsilon}{2}\right] \leq \frac{\delta}{2}, \tag{A.22}$$

and so we need $\omega^{-2} = \frac{\delta}{2}$, and $\omega\mathrm{Var}(X) = \frac{\epsilon}{2}$. Rearranging gives $\mathrm{Var}(X) = \frac{\epsilon\sqrt{\delta}}{2\sqrt{2}}$. Taking ℓ repetitions implies a variance of $\frac{1}{\sqrt{\ell}}$, and so $\ell = \frac{8}{\epsilon^2\delta}$ repetitions suffices to satisfy Eq. A.22.

Now, we consider estimating the X_i's with finite precision, giving an overall estimate $\tilde{X}$, and then bounding the difference between $\tilde{X}$ and X. By Hoeffding's inequality, we have that for n trials:

$$\Pr\left[|\tilde{X} - X| \geq \frac{\epsilon}{2}\right] \leq 2\exp\left\{-\frac{n\epsilon^2}{2}\right\} = \frac{\delta}{2}. \tag{A.23}$$

Rearranging for n, the number of trials, gives

$$n = \frac{2}{\epsilon^2} \log \frac{4}{\delta}. \tag{A.24}$$

The expected number of measurements is

$$n = \sum_{i=1}^{\ell} p_i m_i = \sum_{i=1}^{\ell} \frac{|\langle\psi|\sigma_i|\psi\rangle|^2}{d} m_i. \tag{A.25}$$

It can be readily checked that by substituting in the verifier's choice of m_i, this reduces to Eq.A.24. The guarantee is satisfied by combining Eqs. A.22 and A.23. □

APPENDIX B

TWO-QUBIT FIDELITY ESTIMATION: PROTOCOL RECIPE

We run the minimum-variance fidelity estimation protocol in Thm. 18, for states of the form $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, for a range of different values of θ. However, the photonic chip is only capable of carrying out rank 1 projective measurements; i.e. applying projectors of the form $|\eta\rangle\langle\eta|$ for some product state $|\eta\rangle$. Thus each higher rank projector in the optimal strategy must be "unpacked" into rank 1 components. If we let the total integration time be T, then the optimal protocol is shown in Table B.1, overleaf.

Meas. setting no.	Project onto	Int. time
1	$\|HH\rangle$	$\frac{T}{4}\left(\frac{2-\sin(2\theta)}{4+\sin(2\theta)}\right)$
2	$\|VV\rangle$	$\frac{T}{4}\left(\frac{2-\sin(2\theta)}{4+\sin(2\theta)}\right)$
3	$\|HV\rangle$	$\frac{T}{4}\left(\frac{2-\sin(2\theta)}{4+\sin(2\theta)}\right)$
4	$\|VH\rangle$	$\frac{T}{4}\left(\frac{2-\sin(2\theta)}{4+\sin(2\theta)}\right)$
5	$(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)$	$\frac{T}{3}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
6	$(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
7	$(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
8	$(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{2\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
9	$(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)$	$\frac{T}{3}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
10	$(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
11	$(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\cot\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
12	$(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{4\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{e^{\frac{5\pi i}{3}}}{\sqrt{1+\tan\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
13	$(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{1}{\sqrt{1+\cot\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle-\frac{1}{\sqrt{1+\cot\theta}}\|V\rangle)$	$\frac{T}{3}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
14	$(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{1}{\sqrt{1+\tan\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle-\frac{1}{\sqrt{1+\cot\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
15	$(\frac{1}{\sqrt{1+\tan\theta}}\|H\rangle+\frac{1}{\sqrt{1+\cot\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle+\frac{1}{\sqrt{1+\tan\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$
16	$(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle-\frac{1}{\sqrt{1+\tan\theta}}\|V\rangle)\otimes(\frac{1}{\sqrt{1+\cot\theta}}\|H\rangle+\frac{1}{\sqrt{1+\tan\theta}}\|V\rangle)$	$\frac{T}{9}\left(\frac{1+\sin(2\theta)}{4+\sin(2\theta)}\right)$

Table B.1: The minimum variance fidelity estimation protocol for the state $|\psi\rangle = \sin\theta|00\rangle + \cos\theta|11\rangle$, decomposed into rank 1 projectors. Each block of four settings (1-4, 5-8, 9-12 and 13-16) forms an orthonormal basis; and the integration times sum to T.

APPENDIX C

QUANTUM ALGORITHMS FOR THE FINITE ELEMENT METHOD: ADDITIONAL CALCULATIONS

C.1 Using HHL to approximate the norm of the solution

Assume that we have an s-sparse system of linear equations $A\mathbf{x} = \mathbf{b}$, for some Hermitian $N \times N$ matrix A such that $\lambda_{\max}(A) \leq 1$, $\lambda_{\min}(A) \geq 1/\kappa$. We would like to approximate $\|\mathbf{x}\|$ up to accuracy $\epsilon\|\mathbf{x}\|$ using the HHL algorithm [104]. Here we sketch how the complexity of this task can be bounded, using the same notation as Theorem 31 (see [104] for further technical details). The HHL algorithm is based on a subroutine $\mathcal{P}_{\text{sim}}$ whose probability of acceptance is approximately $p := \|A^{-1}|b\rangle\|^2/\kappa^2$. For any $\delta > 0$, approximating the probability p that a subroutine accepts, up to additive accuracy δp, can be achieved using amplitude estimation [31] with $O(1/(\delta\sqrt{p}))$ uses of the subroutine. Therefore, approximating $\kappa\|\mathbf{b}\|\sqrt{p} = \|\mathbf{x}\|$ up to additive accuracy $\epsilon\|\mathbf{x}\|$ can be achieved with

$$O\left(\frac{\kappa\|\mathbf{b}\|}{\epsilon\|\mathbf{x}\|}\right) = O\left(\frac{\kappa}{\epsilon}\right) \tag{C.1}$$

uses of $\mathcal{P}_{\text{sim}}$, where we use $\lambda_{\max}(A) \leq 1$. The runtime of the $\mathcal{P}_{\text{sim}}$ subroutine, which is described in [104], depends on the accuracy with which its actual probability of acceptance $\widetilde{p}$ approximates p. Using the best known algorithm for Hamiltonian simulation [20] within $\mathcal{P}_{\text{sim}}$, an accuracy of $|\widetilde{p} - p| = O(\epsilon p)$ can be achieved with $O((s\kappa/\epsilon)\,\text{poly}\log(s\kappa/\epsilon))$ uses of the algorithm $\mathcal{P}_A$ for determining entries of A. The runtime is the same up to a polylogarithmic term in N, s, κ, and ϵ. Each use of the subroutine within amplitude estimation requires two uses of $\mathcal{P}_b$ to reflect about the state $|b\rangle$. Therefore, the overall number of uses of $\mathcal{P}_A$ required is

$$O((s\kappa^2/\epsilon)\,\text{poly}\log(s\kappa/\epsilon)), \tag{C.2}$$

and the number of uses of $\mathcal{P}_b$ is $O(\kappa/\epsilon)$. Note that quantum linear equations algorithms subsequent to HHL [6, 52] achieved better dependence on κ, ϵ, or both for the task of producing $|x\rangle$; however, it does not seem obvious how to use these to achieve improved accuracy for estimating $\|\mathbf{x}\|$.

C.2 Bounds on inaccuracies in matrix inversion and classical output

In this appendix we prove the claimed bound in Section 5.6.4 that

$$\frac{\alpha}{\|r\|} = O(h\sqrt{s}), \tag{C.3}$$

where $\alpha = \left(\sum_i \langle \phi_i, r\rangle^2\right)^{1/2}$. Indeed, we show that

$$\sup_{r\neq 0} \frac{\left(\sum_i \langle \phi_i, r\rangle^2\right)^{1/2}}{\|r\|} = O(h\sqrt{s}). \tag{C.4}$$

Observe that this expression will be maximised when r is in the subspace spanned by the $\{\phi_i\}$ functions, so we can assume that $r = \sum_i \mathbf{r}_i \phi_i$ for some $\mathbf{r}_i$. Then the numerator satisfies

$$\begin{aligned}
\left(\sum_i \langle \phi_i, r\rangle^2\right)^{1/2} &= \left(\sum_i \left(\int_\Omega \phi_i(\mathbf{x}) r(\mathbf{x}) d\mathbf{x}\right)^2\right)^{1/2} \\
&= \left(\sum_i \left(\int_\Omega \phi_i(\mathbf{x}) \sum_j \mathbf{r}_j \phi_j(\mathbf{x})\right)^2\right)^{1/2} \\
&= \left(\sum_i \left(\sum_j \mathbf{r}_j \int_\Omega \phi_i(\mathbf{x})\phi_j(\mathbf{x}) d\mathbf{x}\right)^2\right)^{1/2} \\
&= \|W\mathbf{r}\|,
\end{aligned} \tag{C.5}$$

where we define the matrix $W_{ij} := \int_\Omega \phi_i(\mathbf{x})\phi_j(\mathbf{x}) d\mathbf{x}$. Similarly, for the denominator we have

$$\|r\| = \left(\int_\Omega \left(\sum_i \mathbf{r}_i \phi_i(\mathbf{x})\right)^2 d\mathbf{x}\right)^{1/2} = \left(\sum_{i,j} \mathbf{r}_i \mathbf{r}_j \int_\Omega \phi_i(\mathbf{x})\phi_j(\mathbf{x}) d\mathbf{x}\right)^{1/2} = (\mathbf{r}^T W \mathbf{r})^{1/2}. \tag{C.6}$$

Therefore,

$$\frac{\alpha}{\|r\|} \leq \sup_{\mathbf{r}\neq 0} \left(\frac{\mathbf{r}^T W^T W \mathbf{r}}{\mathbf{r}^T W \mathbf{r}}\right)^{1/2} = \sup_{\mathbf{r}', \|\mathbf{r}'\|=1} ((\mathbf{r}')^T W \mathbf{r}')^{1/2} = \|W\|^{1/2}. \tag{C.7}$$

Assume that W is s-sparse. To upper-bound $\|W\|$ we use

$$\|W\| \leq s \max_{i,j} |W_{ij}| = s \max_{i,j} |\langle \phi_i, \phi_j\rangle| \leq s \max_i \|\phi_i\|^2, \tag{C.8}$$

where the first inequality can be found in [51], for example, and the second is Cauchy-Schwarz. Then

$$\|\phi_i\|^2 = \int_T \phi_i(\mathbf{x})^2 d\mathbf{x} \leq h^d \max_{\mathbf{x}\in T} \phi_i(\mathbf{x})^2 = O(h^2), \tag{C.9}$$

where we assume that ϕ_i is supported on a region T of diameter at most h, and we use (5.25) to bound $\max_{\mathbf{x}\in T} \phi_i(\mathbf{x})^2 = O(h^{2-d})$. Thus $\alpha/\|r\| = O(h\sqrt{s})$.

APPENDIX D

SIMULTANEOUS TESTING OF QUANTUM MECHANICS AND GENERAL RELATIVITY WITH A QUANTUM OPTICAL SATELLITE: ADDITIONAL CALCULATIONS

D.1 Derivation of the relativistic time delay

Here we show how to derive Eq. 6.1, generalising the calculation described in [237].

First of all, we identify the reference frame of a distant observer with the ECI (Earth-Centered Inertial) frame. In order to take into account the velocity of the satellite and of the ground station rotating with the Earth's surface, Eq. 7 in [237] must be modified. Let τ_r and t be the proper time recorded by a clock at the radial distance r and the coordinate time recorded by a distant observer, respectively. Then we have:

$$d\tau_r^2 = \left(1 + \frac{2V(r)}{c^2} - \frac{v_r^2}{c^2}\right) dt^2 \tag{D.1}$$

where $V(r) = -\frac{GM}{r}$ is the Earth's gravitational potential and v_r is the speed of the clock at the radial distance r, as measured in the ECI frame.

If the fibre refractive index is n and c is the speed of light in vacuum, the proper time spent by a photon in a fibre loop of proper length l is nl/c. Hence, using Eq. D.1, we can find the time spent by a photon in the two fibres, as measured by a distant observer:

$$t_r = \frac{n l_r}{c\sqrt{1 + \frac{2V(r)}{c^2} - \frac{v_r^2}{c^2}}} \tag{D.2}$$

where l_r is the proper length of the fibre at the radial distance r. Henceforth, the quantities related to the satellite and to the ground station will be indicated by the subscripts s and g, respectively. We assume $l_s = l$ and $l_g = l + dl$.

The difference in photon arrival times (referred to as the "time delay") measured by a distant observer is $t_s - t_g$. Hence, using Eq. D.1 and D.2, the time delay measured by the local observer at the ground station is:

$$\begin{aligned}\Delta\tau &= \sqrt{1+\frac{2V_g}{c^2}-\frac{v_g^2}{c^2}}(t_s-t_g)\\ &= \frac{nl}{c}\left(\frac{\sqrt{1+\frac{V_g}{c^2}-\frac{v_g^2}{c^2}}}{\sqrt{1+\frac{V_s}{c^2}-\frac{v_s^2}{c^2}}}-1\right)-\frac{ndl}{c}.\end{aligned} \tag{D.3}$$

Taking the first order approximation to the potential and the second order approximation to the velocity, Eq. D.3 becomes:

$$\Delta\tau = \frac{nl}{c^3}\left(V_g-\frac{v_g^2}{2}-V_s+\frac{v_s^2}{2}\right)-\frac{ndl}{c}. \tag{D.4}$$

Assuming that the ground station is at rest on the geoid, we can use the conventional geoidal potential W_0 to write the time delay of Eq. D.4 as:

$$\Delta\tau = \frac{nl}{c^3}\left(-W_0-V_s+\frac{v_s^2}{2}\right)-\frac{ndl}{c}. \tag{D.5}$$

Finally, for an elliptical orbit with semi-major axis a, indicating with h the altitude of the satellite on the Earth's surface, the time delay of Eq. D.5 becomes:

$$\Delta\tau = \frac{nl}{c^3}\left[-W_0+GM\left(\frac{2}{R_\oplus+h}-\frac{1}{2a}\right)\right]-\frac{ndl}{c} \tag{D.6}$$

which is exactly Eq. 6.1 in the main chapter.

D.2 Derivation of the expected interferometric signal

Here we show how to obtain Eq. 6.2 in the text.

The commutation relation for continuous mode photon states can be written [144]:

$$[a_{\nu'}, a_\nu^\dagger] = \delta(\nu-\nu'). \tag{D.7}$$

Say we have a Mach-Zehnder interferometer with equal arm lengths, but with one arm subject to time-dilation; then the state at detector $D\pm$ with a single input photon is:

$$|1\rangle_{\nu\pm} \propto \int d\nu f(\nu)\left(e^{i\frac{\nu}{c}(x_r-c\tau_r)} \pm e^{i\frac{\nu}{c}(x_r-c(\tau_r+\Delta\tau))}\right)a_\nu^\dagger|0\rangle.$$

Where x_r is the local Cartesian coordinate (perpendicular to the radial coordinate r), , τ_r is the local time coordinate, $f(\nu)$ is the spectrum of the light source, ν is angular frequency, c is the speed of light, and $\Delta\tau$ is as given in Eq. 6.1.

The probability of a photon arriving at this detector is then

$$\begin{aligned} P_{\pm} =& {}_{v'\pm}\langle 1|1\rangle_{v\pm} \\ &\propto \langle 0|a_{v'}\int dv\int dv' f(v)f(v') \\ &\times\left(e^{i\frac{v}{c}(x_r-c\tau_r)}+e^{i\frac{v}{c}(x_r-c(\tau_r+\Delta\tau)}\right) \\ &\times\left(e^{-i\frac{v'}{c}(x_r-c\tau_r)}+e^{-i\frac{v'}{c}(x_r-c(\tau_r+\Delta\tau)}\right)a_v^{\dagger}|0\rangle \end{aligned} \tag{D.8}$$

From commutation relation Eq. D.7, it is shown that

$${}_{v'}\langle 1|1\rangle_{v} = \delta(v-v') \tag{D.9}$$

Thus

$$\begin{aligned} P_{\pm} \propto& \int dv\int dv'\left(e^{i\frac{v}{c}(x_r-c\tau_r)}+e^{i\frac{v}{c}(x_r-c(\tau_r+\Delta\tau)}\right) \\ &\times f(v)f(v')\left(e^{-i\frac{v'}{c}(x_r-c\tau_r)}+e^{-i\frac{v'}{c}(x_r-c(\tau_r+\Delta\tau)}\right) \\ &\times\delta(v-v') \\ &=\int dv|f(v)\left(e^{i\frac{v}{c}(x_r-c\tau_r)}+e^{i\frac{v}{c}(x_r-c(\tau_r+\Delta\tau)}\right)|^2 \\ &=\frac{1}{2}\left(1+\int dv|f(v)|^2\cos(v\Delta\tau)\right). \end{aligned} \tag{D.10}$$